

## **WO2006022712**

Publication Title:

**METHODS AND AGENTS FOR SCREENING FOR COMPOUNDS CAPABLE OF MODULATING GENE EXPRESSION**

Abstract:

Abstract of WO 2006022712

(A1) The invention relates to the fields of screening assays, compounds, and methods for altering gene expression and protein levels. In particular, the invention includes assays to screen for agents capable of modulating gene expression in a UTR-dependent manner and agents capable of modulating gene expression.

-----

Courtesy of <http://v3.espacenet.com>

(19) World Intellectual Property Organization  
International Bureau



(43) International Publication Date  
2 March 2006 (02.03.2006)

PCT

(10) International Publication Number  
**WO 2006/022712 A1**

(51) International Patent Classification<sup>7</sup>: **C12Q 1/70**,  
1/68, C12N 15/63

**TROTTA, Christopher R.** [US/US]; 198 Scobee Lane,  
Somerset, New Jersey 08873 (US).

(21) International Application Number:  
PCT/US2004/026309

(74) Agents: **MARSH, David R.** et al.; ARNOLD & PORTER  
LLP, 555 Twelfth St., N.W., IP Docketing, Washington,  
District of Columbia 20004 (US).

(22) International Filing Date: 16 August 2004 (16.08.2004)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:  
10/895,393 21 July 2004 (21.07.2004) US

(81) Designated States (*unless otherwise indicated, for every  
kind of national protection available*): AE, AG, AL, AM,  
AT, AU, AZ, BA, BB, BG, BR, BW, BY, BZ, CA, CH, CN,  
CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, EG, ES, FI,  
GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE,  
KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD,  
MG, MK, MN, MW, MX, MZ, NA, NI, NO, NZ, OM, PG,  
PH, PL, PT, RO, RU, SC, SD, SE, SG, SK, SL, SY, TJ, TM,  
TN, TR, TT, TZ, UA, UG, US, UZ, VC, VN, YU, ZA, ZM,  
ZW.

(71) Applicant (*for all designated States except US*): **PTC  
THERAPEUTICS, INC.** [US/US]; 100 Corporate Court,  
Middlesex Business Center, South Plainfield, New Jersey  
07080 (US).

(84) Designated States (*unless otherwise indicated, for every  
kind of regional protection available*): ARIPO (BW, GH,  
GM, KE, LS, MW, MZ, NA, SD, SL, SZ, TZ, UG, ZM,  
ZW), Eurasian (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM),  
European (AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI,  
FR, GB, GR, HU, IE, IT, LU, MC, NL, PL, PT, RO, SE,  
SI, SK, TR), OAPI (BF, BJ, CF, CG, CI, CM, GA, GN,  
GQ, GW, ML, MR, NE, SN, TD, TG).

(72) Inventors; and

(75) Inventors/Applicants (*for US only*): **CAO, Liangx-  
ian** [CN/US]; 6 Mioduski Court, Parlin, New Jersey  
08859 (US). **MEHTA, Anuradha** [IN/US]; 431 Lan-  
caster Court, Piscataway, New Jersey 08854 (US).  
**NARYSHKIN, Nikolai A.** [RU/US]; 137 Windsong  
Circle, East Brunswick, New Jersey 08816 (US). **PELLI-  
GRINI, Matthew C.** [US/US]; 76 Autumn Ridge Road,  
Bedminster, New Jersey 07921 (US). **ROMFO, Charles  
M.** [US/US]; 87 Jonathan Drive, Easton, Pennsylvania  
18045 (US). **TRIFILLIS, Panayiota** [US/US]; 309  
Rivercrest Drive, Piscataway, New Jersey 08854 (US).

Published:

— with international search report

*For two-letter codes and other abbreviations, refer to the "Guid-  
ance Notes on Codes and Abbreviations" appearing at the begin-  
ning of each regular issue of the PCT Gazette.*

(54) Title: METHODS AND AGENTS FOR SCREENING FOR COMPOUNDS CAPABLE OF MODULATING GENE EX-  
PRESSION

(57) Abstract: The invention relates to the fields of screening assays, compounds, and methods for altering gene expression and protein levels. In particular, the invention includes assays to screen for agents capable of modulating gene expression in a UTR-dependent manner and agents capable of modulating gene expression.

WO 2006/022712 A1

## TITLE OF THE INVENTION

Methods and Agents for Screening for Compounds Capable of Modulating Gene Expression

## INCORPORATION OF THE SEQUENCE LISTING

A paper copy of the Sequence Listing and a computer readable form of the sequence listing on diskette, containing the file named "19025.023.SeqList.txt", which is 124,429 bytes in size (measured in MS-DOS), and which was recorded on August 16, 2004, are herein incorporated by reference.

## BACKGROUND OF THE INVENTION

Gene expression, defined as the conversion of the nucleotide sequence of a gene into the nucleotide sequence of a stable RNA or into the amino acid sequence of a protein, is very tightly regulated in every living organism. Regulation of gene expression both of mRNA stability and translation is important in cellular responses to development or environmental stimuli such as nutrient levels, cytokines, hormones, and temperature shifts, as well as environmental stresses like hypoxia, hypocalcemia, viral infection, and tissue injury (reviewed in Guhaniyogi & Brewer, 2001, *Gene* 265(1-2):11-23). Furthermore, alterations in mRNA stability have been causally connected to specific disorders, such as neoplasia, thalassemia, and Alzheimer's disease (reviewed in Guhaniyogi & Brewer, 2001, *Gene* 265(1-2):11-23 and *Translational Control of Gene Expression*, Sonenberg, Hershey, and Mathews, eds., 2000, CSHL Press).

Giordano *et al.*, U.S. Patent No. 6,558,007 (hereafter referred to as "the '007 patent"), assert that they provide a screening assay using a 5' mRNA UTR biased cDNA library or a 3' mRNA UTR biased cDNA library. The '007 patent further asserts that they provide a method of identifying a regulatory UTR sequence using their 5' or 3' mRNA UTR biased cDNA libraries. The '007 patent does not provide assays that mimic the *in vivo* state of a gene controlled by the presence of more than one UTR, for example, genes which are flanked by a 5' UTR and a 3' UTR. Moreover, the approach of the '007 patent requires the libraries described therein.

Pesole *et al.* assert that the 5'- and 3'-UTRs of eukaryotic mRNAs are known to play a crucial role in post-transcriptional regulation of gene expression. Pesole *et al.*, (2002)

*Nucleic Acids Research*, 3(1):335-340, which is hereby incorporated by reference in its entirety. They develop and describe several databases with nucleic acid sequences from UTRs. Many of their database entries are enriched with specialized information including the presence of sequence patterns demonstrated by experimental evidence to play a functional role in gene regulation. Pesole *et al.* do not provide assays to obtain such experimental evidence, nor do they suggest that such experiments mimicked the *in vivo* state of the UTR database entry. Moreover, the methodology of Pesole *et al.* is based on sequence analysis and prior experimental evidence. Pesole *et al.* do not provide experimental screening methods for developing agents to modulate the 5'- and 3'-UTRs of eukaryotic mRNAs that are known to play a crucial role in post-transcriptional regulation of gene expression nor do they suggest a methodology to find novel 5'- and 3'-UTRs of eukaryotic mRNAs that play a crucial role in post-transcriptional regulation of gene expression. In addition, the approach of Pesole *et al.* requires the databases described therein.

Trotta *et al.* assert that the interaction of the La antigen with mdm2 5' UTR enhances mdm2 mRNA translation. Trotta *et al.*, (2003) *Cancer Cell* 3:145-160, which is hereby incorporated by reference in its entirety. They do not suggest methods or agents to screen or identify more UTRs with a similar role in translational regulation of gene expression. Moreover, no agents are provided to screen for compounds that would modulate the regulation of mdm2 mRNA translation.

## SUMMARY OF THE INVENTION

The present invention includes a nucleic acid construct comprising a high-level mammalian expression vector, an intron, and a nucleic acid sequence encoding a reporter polypeptide, wherein said nucleic acid sequence encoding a reporter polypeptide is proximally linked to a target untranslated region (UTR).

The present invention also includes a nucleic acid construct comprising a high-level mammalian expression vector and a nucleic acid sequence encoding a reporter polypeptide, wherein said nucleic acid sequence encoding a reporter polypeptide is directly linked to one or more target UTRs.

The present invention also includes a nucleic acid molecule comprising a nucleic acid sequence encoding a reporter polypeptide directly linked to one or more target UTRs.

The present invention also includes a heterologous population of nucleic acid molecules, wherein said heterologous population comprises a reporter nucleic acid sequence, wherein said nucleic acid sequence encoding a reporter polypeptide is directly linked to one or more target UTRs.

The present invention also includes a method of making a nucleic acid construct to screen for a compound comprising: a) cloning a gene and a vector in said nucleic acid construct; b) engineering said nucleic acid construct to prevent an expressed gene product from having a UTR not found in a target gene; and c) directly linking a target UTR to said gene.

The present invention also includes a method of screening for a compound that modulates expression of a polypeptide comprising: a) maintaining a cell, wherein said cell has a nucleic acid molecule and said nucleic acid molecule comprises a gene encoding a reporter polypeptide and said reporter gene is flanked by a target 5' UTR and a target 3' UTR; b) forming a UTR-complex in said cell; c) contacting a compound with said UTR-complex; and d) detecting an effect of said compound on said UTR-complex.

The present invention also includes a method of screening *in vivo* for a compound that modulates UTR-dependent expression comprising: a) providing a cell having a nucleic acid construct comprising a high-expression, constitutive promoter upstream from a target 5' UTR, said target 5' UTR upstream from a nucleic acid sequence encoding a reporter polypeptide, and said nucleic acid sequence encoding a reporter polypeptide upstream from a target 3' UTR; b) contacting said cell with a compound; c) producing a nucleic acid molecule that contains a nucleic acid sequence encoding a reporter polypeptide and does not contain UTR not found in a target gene; and d) detecting said reporter polypeptide.

The present invention also includes a method of screening *in vitro* for a compound that modulates UTR-affected expression comprising: a) providing an *in vitro* translation system; b) contacting said *in vitro* translation system with a compound and a nucleic acid molecule comprising a target 5' UTR, said target 5' UTR upstream from a nucleic acid sequence encoding a reporter polypeptide and said nucleic acid sequence encoding a reporter polypeptide upstream from a target 3' UTR, wherein said nucleic acid molecule is in an absence of a UTR not found in a target gene; and c) detecting said reporter polypeptide *in vitro*.

The present invention also includes a method of expressing a nucleic acid molecule in a cell comprising: a) providing a heterologous nucleic acid molecule to a cell, wherein said nucleic acid molecule comprises a nucleic acid sequence encoding a reporter polypeptide flanked by target UTRs in an absence of a UTR not found in a target gene; and b) detecting said reporter polypeptide *in vivo*.

The present invention also includes a method of screening for a compound that modulates protein expression through a main ORF-independent, UTR-affected mechanism comprising: a) growing a stable cell line having a reporter gene proximally linked to a target UTR; b) comparing said stable cell line in the presence of a compound relative to in an absence of said compound; and c) selecting for said compound that modulates protein expression through a main ORF-independent, UTR-affected mechanism.

The present invention also includes a method of screening for a compound that modulates protein expression through a main ORF-independent, UTR-affected mechanism comprising: a) substituting in a cell a target gene with a reporter gene, wherein proximally linked target UTRs of said target gene remain intact and said cell is a differentiated cell; b) growing said cell line; and c) selecting for said compound that modulates protein expression of said reporter gene through a main ORF-independent, UTR-affected mechanism.

The present invention also includes a method of screening for a compound that modulates protein expression through a UTR-affected mechanism comprising: a) growing a stable cell line having a reporter gene proximally linked to a target UTR, wherein said stable cell line mimics post-transcriptional regulation of a target gene found *in vivo*; b) growing said stable cell line; and c) selecting for said compound that modulates protein expression of said reporter gene through a UTR-affected mechanism.

The present invention also includes a method of screening for a compound that modulates protein expression through a UTR-affected mechanism comprising: a) growing a stable cell line having a reporter gene proximally linked to more than one target UTR; b) comparing said stable cell line in the presence of a compound relative to in an absence of said compound, wherein said compound does not modulate UTR-dependent expression if only one target UTR is proximally linked to a reporter gene; and c) selecting for said compound that modulates protein expression of said reporter gene through a UTR-affected mechanism.

## BRIEF DESCRIPTION OF THE DRAWINGS

Figure 1 sets forth the UTR specificity for reporter gene expression when flanked by the HIF 1 $\alpha$  5' and 3' UTR (5+3 UTR). The reporter gene is operably linked to the HIF 1 $\alpha$  5' and 3' UTR (5+3 UTR), the HIF 1 $\alpha$  5' UTR (5' UTR), the HIF 1 $\alpha$  3' UTR (3' UTR), or no HIF 1 $\alpha$  UTR (No UTR) under conditions of normoxia and hypoxia.

Figure 2A sets forth a schematic of the construct indicating the locations of primers used.

Figure 2B sets forth the results of RT-PCR described in Example 2 to determine the quality of stable clones using the primers indicated in Figure 2A.

Figure 2C sets forth the results of RT-PCR described in Example 2 to determine the quality of stable clones using the primers indicated in Figure 2A.

Figure 3 sets forth the luciferase activity per microgram of total protein for each stable clonal cell line indicated.

Figure 4 sets forth the luciferase activity per microgram of total protein as a function of the fold increase over the level of actin RNA.

Description of the Nucleic Acid Sequences

SEQ ID NO: 1 sets forth a luciferase 5' reverse primer.

SEQ ID NO: 2 sets forth a luciferase 3' forward primer.

SEQ ID NO: 3 sets forth a FLuc F.

SEQ ID NO: 4 sets forth a FLuc R.

SEQ ID NO: 5 sets forth a FLuc probe.

SEQ ID NO: 6 sets forth a homo sapiens VEGF 5' UTR, derived from Accession No. NM\_03376 of AF095785.

SEQ ID NO: 7 sets forth a 3' UTR is derived from Accession No. AF022375, genomic contig where sequences are derived is VEGF - NT\_007592.

SEQ ID NO: 8 sets forth a homo sapiens TNF-alpha 5' UTR derived from Accession No. NM\_00594.

SEQ ID NO: 9 sets forth a homo sapiens TNF-alpha 3' UTR derived from Accession No. NM\_00594.

SEQ ID NO: 10 sets forth an ARE 1 from homo sapiens TNF-alpha 3' UTR derived from Accession No. NM\_00594.

SEQ ID NO: 11 sets forth an ARE 1 from homo sapiens TNF-alpha 3' UTR derived from Accession No. NM\_00594.

SEQ ID NO: 12 sets forth an ARE 1 from homo sapiens TNF-alpha 3' UTR derived from Accession No. NM\_00594.

SEQ ID NO: 13 sets forth a constitutive decay element (hereinafter "CDE") derived from homo sapiens TNF-alpha 3' UTR as discussed in Stoecklin *et al.*, (2003) *Molecular and Cellular Biology*, 23(10):3506-3515, which is hereby incorporated by reference in its entirety.

SEQ ID NO: 14 sets forth a putative second ARE from homo sapiens TNF-alpha 3' UTR derived from Accession No. NM\_00594.

SEQ ID NO: 15 sets forth a putative poly(A) signal from homo sapiens TNF-alpha 3' UTR derived from Accession No. NM\_00594.

SEQ ID NO: 16 sets forth a homo sapiens MDM2 5' UTR as derived from Accession No. NM\_002392.

SEQ ID NO: 17 sets forth a homo sapiens Her-2 5' UTR sequence derived from Accession No. NM\_004448.

SEQ ID NO: 18 sets forth a homo sapiens Her-2 5' uORF sequence derived from Accession No. NM\_004448.

SEQ ID NO: 19 sets forth a Her-2 3' UTR derived from Accession No. NM\_004448.

SEQ ID NO: 20 sets forth a 336 nucleotide region of a VEGF 5' UTR

SEQ ID NO: 21 sets forth a 476 nucleotide region of a VEGF 5' UTR.

SEQ ID NO: 22 sets forth a 73 nucleotide sequence from a Her-2 3' UTR.

SEQ ID NO: 23 sets forth a 81 nucleotide region native to pcDNA<sup>TM</sup>3.1/Hygro (Invitrogen Corp., Carlsbad, CA).

SEQ ID NO: 24 sets forth a 134 nucleotide region native to pcDNA<sup>TM</sup>3.1/Hygro (Invitrogen Corp., Carlsbad, CA).

### Definitions

As used herein, the term "construct" refers to an artificially manipulated nucleic acid molecule.

As used herein, the term "gene" is a segment of DNA that is capable of producing a polypeptide.



As used herein, the term “heterologous” refers to ingredients or constituents of dissimilar or diverse origin.

As used herein, the term “mammalian cancer cell” or “mammalian tumor cell” refers to a cell derived from a mammal that proliferates inappropriately.

As used herein, the term “main ORF-independent mechanism” refers to a cellular pathway or process, wherein at least one step relates to gene expression and is not dependent on the nucleic acid sequence of the main open reading frame.

As used herein, the term “reporter gene” refers to any gene whose expression can be measured.

As used herein, the term “RNA induced gene silencing, or RNA interference (RNAi)” refers to the mechanism of double-stranded RNA (dsRNA) introduced into a system to reduce protein expression of specific genetic sequence.

As used herein, the term “specifically bind” means that a compound binds to another compound in a manner different from a similar type of compounds, e.g. in terms of affinity, avidity, and the like. In a non-limiting example, more binding occurs in the presence of a competing reagent, such as casein. In another non-limiting example, antibodies that specifically bind a target protein should provide a detection signal at least 2-, 5-, 10-, or 20-fold higher relative to a detection signal provided with other molecules when used in Western blots or other immunochemical assays. In an alternative non-limiting example, a nucleic acid can specifically bind its complementary nucleic acid molecule. In another non-limiting example, a transcription factor can specifically bind a particular nucleic acid sequence.

As used herein, the term “secondary structure” means the alpha-helical, beta-sheet, random coil, beta turn structures and helical nucleic acid structures that occur in proteins, polypeptides, nucleic acids, compounds comprising modified nucleic acids, compounds comprising modified amino acids and other types of compounds as a result of, at least, the compound's composition.

As used herein, the term “non-peptide therapeutic agent” and analogous terms include, but are not limited to organic or inorganic compounds (*i.e.*, including heteroorganic and organometallic compounds but excluding proteins, polypeptides and nucleic acids).

As used herein, the term “uORF” refers to an upstream open reading frame that is in the 5' UTR of the main open reading frame, *i.e.*, that encodes a functional protein, of a mRNA.

As used herein, the term “UTR” refers to the untranslated region of a mRNA.

As used herein, the term “untranslated region-dependent expression” or “UTR-dependent expression” refers to the regulation of gene expression through UTRs at the level of mRNA expression, *i.e.*, after transcription of the gene has begun until the protein or the RNA product(s) encoded by the gene has been degraded or excreted.

As used herein, the term “vector” refers to a nucleic acid molecule used to introduce a nucleic acid sequence in a cell or organism.

## DETAILED DESCRIPTION OF THE INVENTION

The present invention includes and utilizes the fact that an untranslated region (UTR) is capable of modulating expression of a gene and that such modulation of expression is capable of being altered or modulated by the addition of compounds. In a preferred embodiment, a UTR is a region of a RNA that is not translated into protein. In a more preferred embodiment, a UTR is a flanking region of the RNA transcript that is not translated into the targeted protein, and can include a 5' UTR that has a short, putative open reading frame. In a most preferred embodiment, the UTR is a 5' UTR, *i.e.*, upstream of the coding region, or a 3' UTR, *i.e.*, downstream of the coding region.

Moreover, the present invention includes and provides agents and methods useful in screening for a compound capable of modulating gene expression and also hybrid molecules.

### Nucleic Acid Agents and Constructs

One skilled in the art may refer to general reference texts for detailed descriptions of known techniques discussed herein or equivalent techniques. These texts include Ausubel *et al.*, *Current Protocols in Molecular Biology*, John Wiley and Sons, Inc. (1995); Sambrook *et al.*, *Molecular Cloning, A Laboratory Manual* (2d ed.), Cold Spring Harbor Press, Cold Spring Harbor, New York (1989); Birren *et al.*, *Genome Analysis: A Laboratory Manual*, volumes 1 through 4, Cold Spring Harbor Press, Cold Spring Harbor, New York (1997-

1999). These texts can, of course, also be referred to in making or using an aspect of the invention.

#### *UTRs*

The present invention includes nucleic acid molecules with UTRs that comprise or consist of a gene expression modulator (GEM), fragments thereof, and complements of each. As used herein, a UTR can be a naturally occurring genomic DNA sequence. In a preferred embodiment, a UTR is a 5' UTR, *i.e.*, upstream of the coding region, or a 3' UTR, *i.e.*, downstream of the coding region.

In one embodiment, a UTR of the present invention comprises or consists of a nucleic acid sequence selected from a group consisting of SEQ ID NOs: 6-22, and including fragments of each, and complements of all. In another embodiment, a nucleic acid molecule of the present invention contains or comprises a nucleic acid sequence that is greater than 85% identical, and more preferably greater than 86, 87, 88, 89, 90, 91, 92, 93, 94, 95, 96, 97, 98, or 99% identical to a UTR of the present invention, a GEM nucleic acid sequence, a complement of either, and a fragment of any of these sequences.

The hybridization conditions typically involve nucleic acid hybridization in about 0.1X to about 10X SSC (diluted from a 20X SSC stock solution containing 3 M sodium chloride and 0.3M sodium citrate, pH 7.0 in distilled water), about 2.5X to about 5X Denhardt's solution (diluted from a 50X stock solution containing 1% (w/v) bovine serum albumin, 1% (w/v) Ficoll<sup>®</sup> (Amersham Biosciences Inc., Piscataway, NJ), and 1% (w/v) polyvinylpyrrolidone in distilled water), about 10 mg/ml to about 100 mg/ml salmon sperm DNA, and about 0.02% (w/v) to about 0.1% (w/v) SDS, with an incubation at about 20° C to about 70° C for several hours to overnight.

In a preferred aspect, the moderate stringency hybridization conditions are provided by 6X SSC, 5X Denhardt's solution, 100 mg/ml salmon sperm DNA, and 0.1% (w/v) SDS, with an incubation at 55° C for several hours. The moderate stringency wash conditions are about 0.02% (w/v) SDS, with an incubation at about 55° C overnight. In a more preferred aspect, the high stringency hybridization conditions are about 2X SSC, about 3X Denhardt's solution, and about 10 mg/ml salmon sperm DNA. The high stringency wash conditions are about 0.05% (w/v) SDS, with an incubation at about 65° C overnight.

The percent identity is preferably determined using the “Best Fit” or “Gap” program of the Sequence Analysis Software Package™ (Version 10; Genetics Computer Group, Inc., University of Wisconsin Biotechnology Center, Madison, WI). “Gap” utilizes the algorithm of Needleman and Wunsch to find the alignment of two sequences that maximizes the number of matches and minimizes the number of gaps. “BestFit” performs an optimal alignment of the best segment of similarity between two sequences and inserts gaps to maximize the number of matches using the local homology algorithm of Smith and Waterman. The percent identity calculations may also be performed using the Megalign program of the LASERGENE bioinformatics computing suite (default parameters, DNASTAR Inc., Madison, Wisconsin). The percent identity is most preferably determined using the “Best Fit” program using default parameters.

Any of a variety of methods may be used to obtain one or more of the above-described nucleic acid molecules of the present invention. Automated nucleic acid synthesizers may be employed for this purpose. In lieu of such synthesis, the disclosed nucleic acid molecules may be used to define a pair of primers that can be used with the polymerase chain reaction (PCR) to amplify and obtain any desired nucleic acid molecule or fragment.

Short nucleic acid sequences having the ability to specifically hybridize to complementary nucleic acid sequences may be produced and utilized in the present invention, *e.g.*, as probes to identify the presence of a complementary nucleic acid sequence in a given sample. Alternatively, the short nucleic acid sequences may be used as oligonucleotide primers to amplify or mutate a complementary nucleic acid sequence using PCR technology. These primers may also facilitate the amplification of related complementary nucleic acid sequences (*e.g.*, related sequences from other species).

Use of these probes or primers may greatly facilitate the identification of transgenic cells or organisms that contain the presently disclosed structural nucleic acid sequences. Such probes or primers may also, for example, be used to screen cDNA, mRNA, or genomic DNA libraries for additional nucleic acid sequences related to or sharing homology with the presently disclosed promoters and structural nucleic acid sequences. The probes may also be PCR probes, which are nucleic acid molecules capable of initiating a polymerase activity while in a double-stranded structure with another nucleic acid.

A primer or probe is generally complementary to a portion of a nucleic acid sequence that is to be identified, amplified, or mutated and of sufficient length to form a stable and sequence-specific duplex molecule with its complement. The primer or probe preferably is about 10 to about 200 residues long, more preferably is about 10 to about 100 residues long, even more preferably is about 10 to about 50 residues long, and most preferably is about 14 to about 30 residues long.

The primer or probe may, for example without limitation, be prepared by direct chemical synthesis, by PCR (U.S. Patent Nos. 4,683,195 and 4,683,202), or by excising the nucleic acid specific fragment from a larger nucleic acid molecule. Various methods for determining the sequence of PCR probes and PCR techniques exist in the art. Computer-generated searches using programs such as Primer3 ([www-genome.wi.mit.edu/cgi-bin/primer/primer3.cgi](http://www-genome.wi.mit.edu/cgi-bin/primer/primer3.cgi)), STSPipeline ([www-genome.wi.mit.edu/cgi-bin/www-STSPipeline](http://www-genome.wi.mit.edu/cgi-bin/www-STSPipeline)), or GeneUp (Pesole *et al.*, *BioTechniques* 25:112-123, 1998), for example, can be used to identify potential PCR primers.

Furthermore, sequence comparisons can be done to find nucleic acid molecules of the present invention based on secondary structure homology. Several methods and programs are available to predict and compare secondary structures of nucleic acid molecules, for example, GeneBee (available on the world wide web at [genebee.msu.su/services/rna2\\_reduced.html](http://genebee.msu.su/services/rna2_reduced.html)); the Vienna RNA Package (available on the world wide web at [tbi.univie.ac.at/~ivo/RNA/](http://tbi.univie.ac.at/~ivo/RNA/)); SstructView (available on the world wide web at the Stanford Medical Informatics website, under: [projects/helix/sstructview/home.html](http://projects/helix/sstructview/home.html) and described in "RNA Secondary Structure as a Reusable Interface to Biological Information Resources." 1997. *Gene* vol. 190GC59-70). For example, comparisons of secondary structure are preformed in Le *et al.*, A common RNA structural motif involved in the internal initiation of translation of cellular mRNAs. 1997. *Nuc. Acid. Res.* vol. 25(2):362-369.

#### *UTR-complexes*

The present invention also includes a UTR that is complexed. A UTR-complex includes a complex of two or more identical UTRs, one or more different UTRs, a pair of UTRs from the same gene, one or more UTRs and one or more proteins, one or more UTRs and one or more nucleic acids, one or more UTRs and one or more nucleic acid molecules.

By way of non-limiting examples, a UTR-complex can be a complex of a UTR and a small interfering RNAs (siRNA), a UTR and a RNA/DNA sense strand, or a UTR and a RNA/DNA antisense strand.

A UTR-complex of the present invention can refer to a non-covalent or covalent attachment to a UTR. In a preferred embodiment, a GEM or UTR of a nucleic acid molecule modulates attachment of complex constituents to the nucleic acid molecule that has a UTR. In a more preferred embodiment, a UTR-complex varies depending on the nucleic acid sequence of the UTR within the nucleic acid molecule. In a most preferred embodiment, the nucleic acid sequence of the UTR that affects a UTR-complex indicates the presence of a GEM. In a preferred embodiment, the UTR, a GEM, or a fragment of either, modulates the formation of a UTR-complex. In an alternate embodiment, the UTR, or a fragment thereof, modulates the disassociation, the stability, or the constituents of the UTR-complex. In a preferred embodiment, the non-covalent or covalent attachment is a transient attachment. In a more preferred embodiment, the constituents of a UTR-complex vary during processing. In a most preferred embodiment, the constituents of the UTR-complex vary depending on the nucleic acid sequence of the UTR within the nucleic acid molecule, which is in the presence of cellular proteins that can be cell-type specific.

A UTR-complex of the present invention can include the non-covalent or covalent attachment of one or more ribonucleoproteins to a nucleic acid molecule that contains a UTR. In a preferred embodiment, a GEM of the present invention or a UTR of a nucleic acid molecule of the present invention modulates the attachment of the nucleic acid molecule and one or more ribonucleoproteins.

By way of non-limiting examples, UTR-complexes are provided in Pesole *et al.* and Trotta *et al.*, cited and incorporated by reference above, as well as on the world wide web, including at the ftp site: [bighost.ba.itb.cnr.it/pub/Embnet/Database/UTR/](http://bighost.ba.itb.cnr.it/pub/Embnet/Database/UTR/) (as available on July 20, 2004), which is hereby incorporated by reference in its entirety. Furthermore, a GEM or UTR of the present invention can interact with a protein from the large family of AU-rich containing mRNAs associated with Hu-Antigen R (HuR)-mediated regulation (including IL-3, c-fos, c-myc, GM-CSF, AT-R1, Cox-2, IL-8 or TNF- $\alpha$  as cited in WO 03/087815), the RNA recognition motif (RRM) superfamily, the small nuclear RNPs (snRNP), hnRNP proteins, mRNA proteins, exon junction complex (EJC) proteins,

cytoplasmic exon junction complex (cEJC) proteins, U snRNA proteins, nuclear pore complex proteins, dead-box family proteins, splicing factors, ribosomal proteins, and translation-specific proteins that are non-ribosomal, non-regulatory ribosomal protein, and chromatin-associated protein. For specific examples *see* Dreyfuss, *et al.* (2002) *Nature Reviews: Molecular Cell Biology* 3:195-205, hereby incorporated in its entirety. See also on the world wide web at the ftp site: [ftp.ebi.ac.uk/pub/databases/UTR/](ftp://ftp.ebi.ac.uk/pub/databases/UTR/) (as available on July 21, 2004), which is hereby incorporated by reference in its entirety. In the present invention, splicing factors include, but are not limited to, serine-arginine (SR) proteins. In the present invention, translation-specific proteins that are non-ribosomal include, without limitation, exon-junction complex proteins, poly-A binding proteins, and cap-binding proteins.

Other examples of UTR-complexes include a TNF- $\alpha$  mRNA complexed with the tristetraprolin protein (TTP; *see* Lai *et al.*, (1999) *Molecular and Cellular Biology*, 19(6):4311-4323, hereby incorporated by reference in its entirety) and TIA-1 bound to AREs in the 3' UTRs. The TIA-1 recognition results in more TIAs binding to the first TIA-1. This TIA complex recognizes the 40 S ribosome subunit which is bound to the 5' UTR. Therefore, preventing the TIA-1 from binding to the AREs prevents translation of the encoded protein upstream of the bound ARE in the 3' UTR. *See* Kedersha and Anderson, (2002) *Biochemical Society Transactions*, 30(6):963-969, hereby incorporated by reference in its entirety.

#### Constructs of the Present Invention

The present invention includes and provides nucleic acid constructs. It is understood that any of the constructs and other nucleic acid agents of the present invention can be either DNA or RNA. In a preferred embodiment, a construct can be a nucleic acid molecule having a UTR, a coding sequence, or both. In another embodiment, a construct is composed of at least one UTR of the present invention, a sequence encoding a reporter polypeptide, and a vector. Moreover, any of the nucleic acid molecules of the present invention can be used in combination with a method of the present invention.

#### *Vectors*

Exogenous genetic material may be introduced into a host cell by use of a vector or construct designed for such purpose. Any of the nucleic acid sequences of the present invention can be incorporated into a vector or construct of the present invention. A vector or

construct of the present invention includes, without limitation, linear or closed circular plasmids. A vector system may be a single vector or plasmid or two or more vectors or plasmids that together contain the total DNA to be introduced into the genome of the host. In a preferred embodiment, a vector contains a promoter functional in mammalian cells or bacteria or both. Methods for preparing vectors or constructs are well known in the art.

Vectors suitable for replication in mammalian cells may include viral replicons, or sequences that insure integration of the appropriate sequences encoding HCV epitopes into the host genome. For example, another vector used to express foreign DNA is vaccinia virus. Such heterologous DNA is generally inserted into a gene that is non-essential to the virus, for example, the thymidine kinase gene (tk), which also provides a selectable marker. Expression of the HCV polypeptide then occurs in cells or animals that are infected with the live recombinant vaccinia virus.

In general, plasmid vectors containing replicon and control sequences that are derived from species compatible with the host cell are used in connection with bacterial hosts. The vector ordinarily carries a replication site, as well as marking sequences that are capable of providing phenotypic selection in transformed cells. For example, *E. coli* is typically transformed using a construct with a backbone derived from a vector, such as pBR322, which contains genes for ampicillin and tetracycline resistance and thus provides easy approach for identifying transformed cells. The pBR322 plasmid, or other microbial plasmid or phage, also generally contains, or is modified to contain, promoters that can be used by the microbial organism for expression of the selectable marker genes.

In a preferred embodiment of the present invention, an expression vector can be a high-level mammalian expression vector designed to randomly integrate into the genome, for example, pCMR1. A high-level expression vector will have about 100 to about 1000 copies per cell, about 100 to about 500 copies per cell, about 500 to about 1000 copies per cell, or about 250 to about 1000 copies per cell. In one embodiment, a high-level mammalian expression vector is derived from the family of pUC vectors. In a preferred embodiment of the present invention, an expression vector can be a high-level mammalian expression vector designed to site-specifically integrate into the genome of cells. For example, pMCP1 can site-specifically integrate into the genome of cells genetically engineered to contain the FRT



site-specific recombination site via the Flp recombinase (see, *e.g.*, Craig, 1988, *Ann. Rev. Genet.* 22: 77-105; and Sauer, 1994, *Curr. Opin. Biotechnol.* 5: 521-527).

#### *Promoters*

A construct can include a promoter, *e.g.*, a recombinant vector typically comprises, in a 5' to 3' orientation: a promoter to direct the transcription of a nucleic acid molecule of interest.

In a preferred aspect of the present invention, a construct can include a mammalian promoter and can be used to express a nucleic acid molecule of choice. As used herein, a "mammalian promoter" refers to a promoter functional in a mammalian cell, derived from a mammalian cell, or both. A number of promoters that are active in mammalian cells have been described in the literature. A promoter can be selected on the basis of the cell type into which the vector will be inserted.

A preferred promoter of the present invention is an endogenous promoter. A particularly preferred promoter is upstream from the target gene that has its expression modulated by a GEM. Other promoter sequences can be utilized in a construct or other nucleic acid molecules, suitable promoters include, but are not limited to, those described herein.

Suitable promoters for mammalian cells are known in the art and include viral promoters, such as those from Simian Virus 40 (SV40), Rous sarcoma virus (RSV), adenovirus (ADV), cytomegalovirus (CMV), and bovine papilloma virus (BPV) as well as the parvovirus B19p6 promoter and mammalian cell-derived promoters. A number of viral-based expression systems can be used to express a reporter gene in mammalian host cells. For example, if an adenovirus is used as an expression vector, sequences encoding a reporter gene can be ligated into an adenovirus transcription/translation complex comprising the late promoter and tripartite leader sequence.

Other examples of preferred promoters include tissue-specific promoters and inducible promoters. Other preferred promoters include the hematopoietic stem cell-specific, *e.g.*, CD34, glucose-6-phosphatase, interleukin-1 alpha, CD11c integrin gene, GM-CSF, interleukin-5R alpha, interleukin-2, c-fos, h-ras and DMD gene promoters. Other promoters include the herpes thymidine kinase promoter, and the regulatory sequences of the metallothionein gene.

Inducible promoters suitable for use with bacteria hosts include the  $\beta$ -lactamase and lactose promoter systems, the arabinose promoter system, alkaline phosphatase, a tryptophan (trp) promoter system and hybrid promoters such as the tac promoter. However, other known bacterial inducible promoters are suitable. Promoters for use in bacterial systems also generally contain a Shine-Dalgarno sequence operably linked to the DNA encoding the polypeptide of interest.

A promoter can also be selected on the basis of their regulatory features, *e.g.*, enhancement of transcriptional activity, inducibility, tissue specificity, and developmental stage-specificity. A promoter can work *in vitro*, for example the T7-promoter. Particularly preferred promoters can also be used to express a nucleic acid molecule of the present invention in a nonhuman mammal. Additional promoters that may be utilized are described, for example, in Benoist and Chambon, *Nature* 290:304-310 (1981); Yamamoto *et al.*, *Cell* 22:787-797 (1980); Wagner *et al.*, *PNAS* 78:1441-1445 (1981); Brinster *et al.*, *Nature* 296:39-42 (1982).

#### *Main ORF*

Agents and constructs of the invention can include nucleic acid molecules with a main ORF. As used herein, a "main ORF" is a nucleic acid sequence, including sequence in deoxyribonucleic acid or ribonucleic acid molecules, that codes for a polypeptide. As used herein, the term "main ORF DNA" refers to the open reading frame of a gene, *i.e.*, the region of the gene that is translated into protein. As used herein, the term "ORF" refers to the open reading frame of a mRNA, *i.e.*, the region of the mRNA that is translated into protein. In a preferred embodiment of the present invention, a main ORF can be in a gene with an upstream open reading frame ("uORF") contained in the 5' UTR of the gene. As used herein, the term "uORF" refers to an upstream open reading frame that is in the 5' UTR of the main open reading frame, *i.e.*, that encodes a functional protein, of a mRNA.

As used herein, a "control gene" can be any gene that is not identical to a target gene being used. In a preferred embodiment, a control gene is a gene that does not contain a GEM. In a most preferred embodiment, a control gene is a target gene with GEM sequence removed or altered to be ineffective.

#### Target genes

As used herein, the term “target gene” refers to a gene or nucleotide sequence encoding a protein or polypeptide of interest. In a preferred embodiment, target genes are selected for investigation based on 1) role of a target gene in a disease phenotype; 2) post-transcriptional control of a target gene’s expression; and 3) commercial considerations, including but not limited to medical need, market size, and competition.

In a highly preferred embodiment, a target gene can be myostatin, utrophin, alpha 7 integrin, insulin like growth factor 1, or phospholamban. In a most preferred embodiment, a target gene can be utrophin isoform A, alpha 7 integrin isoforms X2A, X2DA, X2B, and X2DB (which are muscle specific), insulin like growth factor 1 isoform exon1-Ea expressed in extrahepatic tissues, or insulin like growth factor 1 isoform exon1-MGF expressed specifically in skeletal muscle.

In a preferred embodiment, target genes are selected from the group of target genes with a role in a disease or condition including, but not limited to, skin disease, cancer, inflammatory diseases, asthma, rheumatoid arthritis, multiple sclerosis (MS), Alzheimer’s disease, autoimmunity, systemic lupus erythematosus (SLE), Crohn’s disease, genetic diseases, diabetes, obesity, neurologic disease, central nervous system (CNS) diseases, Parkinson’s disease, pain response abnormality, schizophrenia, Huntington’s disease, cardiovascular disease, anti-infective diseases, human immune deficiency (HIV), hepatitis C virus (HCV), hepatitis B virus (HBV), hepatitis A virus (HAV), and cholera.

Particularly preferred target genes can have a role in more than one disease, including, but not limited to, combinations such as cancer and inflammatory diseases; inflammatory diseases and asthma, rheumatoid arthritis, multiple sclerosis, Alzheimer’s disease, autoimmunity, SLE, Crohn’s disease, or combinations of any or all of these; diabetes and obesity; diabetes and neurologic disease; CNS and Alzheimer’s disease, pain response abnormality, Parkinson’s disease, Huntington’s disease, schizophrenia, anti-infective diseases and inflammatory diseases, cancer, HIV, HCV, HBV, HAV, cholera, or combinations of any or all of these; and combinations of these disease combinations.

In a most preferred embodiment, target genes have specific functions in promoting the disease or condition, such as, but not limited to, enzymes of sugar metabolism, involved in glucose homeostasis control, and involved in satiety and weight control. In a preferred

embodiment, target genes do not include bovine growth factor hormone, adenaline repeats, reporter sequences, or epitope tags like myc or HLA.

### Reporter genes

As used herein, a “reporter gene” is any gene whose expression can be measured. In a preferred embodiment, a reporter gene does not have any UTRs. In a more preferred embodiment, a reporter gene is a contiguous open reading frame. In another preferred embodiment, a reporter gene can have a previously determined reference range of detectable expression.

Constructs of the invention can comprise one or more reporter genes fused to one or more UTRs. For example, specific RNA sequences, RNA structural motifs, and/or RNA structural elements that are known or suspected to modulate UTR-dependent expression of a target gene can be fused to the reporter gene. A reporter gene of the present invention encoding a protein, a fragment thereof, or a polypeptide, can also be linked to a propeptide encoding region. A propeptide is an amino acid sequence found at the amino terminus of a proprotein or proenzyme. The resulting polypeptide is known as a propolypeptide or proenzyme (a zymogen in some cases). Propolypeptides are generally inactive and can be converted to mature active polypeptides by catalytic or autocatalytic cleavage of the propeptide from the propolypeptide or proenzyme.

A reporter gene can express a selectable or screenable marker. Selectable markers may also be used to select for organisms or cells that contain exogenous genetic material. Examples of such include, but are not limited to: a *neo* gene (which codes for kanamycin resistance and can be selected for using kanamycin), GUS, green fluorescent protein (GFP), neomycin phosphotransferase II (*nptII*), luciferase (LUX), or an antibiotic resistance coding sequence. Screenable markers can be used to monitor expression. Exemplary screenable markers include: a  $\beta$ -glucuronidase or *uidA* gene (GUS) which encodes an enzyme for which various chromogenic substrates are known; a  $\beta$ -lactamase gene, a gene which encodes an enzyme for which various chromogenic substrates are known (*e.g.*, PADAC, a chromogenic cephalosporin); a luciferase gene; a tyrosinase gene, which encodes an enzyme capable of oxidizing tyrosine to DOPA and dopaquinone which in turn condenses to melanin; and  $\alpha$ -galactosidase, which can be used in colormetric assays.

Included within the terms “selectable or screenable marker genes” are also genes that encode a secretable marker whose secretion can be detected as a method of identifying or selecting for transformed cells. Examples include markers that encode a secretable antigen that can be identified by antibody interaction, or even secretable enzymes, which can be detected utilizing their inherent biochemical properties. Secretable proteins fall into a number of classes, including small, diffusible proteins which are detectable, (*e.g.*, by ELISA), or small active enzymes which are detectable in extracellular solution (*e.g.*,  $\alpha$ -amylase,  $\beta$ -lactamase, phosphinothricin transferase). Other possible selectable or screenable marker genes, or both, are apparent to those of skill in the art.

A reporter gene can express a fusion protein. As such, the fusion protein can be a fusion of any reporter gene operably linked to another gene, or fragment thereof. For instance, the expressed fusion protein can provide a “tagged” epitope to facilitate detection of the fusion protein, such as GST, GFP, FLAG, or polyHIS. Such fusions preferably encode between 1 and 50 amino acids, more preferably between 5 and 30 additional amino acids, and even more preferably between 5 and 20 amino acids. In one embodiment, a fusion protein can be a fusion protein that includes in whole or in part of a target protein sequence.

Alternatively, the fusion can provide regulatory, enzymatic, cell signaling, or intercellular transport functions. For example, a sequence encoding a signal peptide can be added to direct a fusion protein to a particular organelle within a eukaryotic cell. Such fusion partners preferably encode between 1 and 1000 additional amino acids, more preferably between 5 and 500 additional amino acids, and even more preferably between 10 and 250 amino acids.

In one embodiment, a reporter gene includes one or more mutations (*e.g.*, one or more substitutions, deletions and/or additions) that do not alter the ability of reporter gene expression to be measured. In a highly preferred embodiment, the reporter gene contains one or more restriction sites that can be used for cloning, such as a *Bam*HI and a *Not* I site, and the restriction sites do not alter the function of the reporter gene. In a particularly preferred embodiment, a restriction site is downstream from the start codon of the open reading frame that encodes the reporter polypeptide, and another restriction site is upstream from the stop codon of the open reading frame that encodes the reporter polypeptide.

The present invention also provides for a reporter gene flanked by one or more

untranslated regions (*e.g.*, the 5' UTR, 3' UTR, or both the 5' UTR and 3' UTR of the target gene). In addition, the present invention provides for a reporter gene flanked by one or more UTRs of a target gene, where the UTR contains one or more mutations (*e.g.*, one or more substitutions, deletions and/or additions). In a preferred embodiment, the reporter gene is flanked by both 5' and 3' UTRs so that compounds that interfere with an interaction between the 5' and 3' UTRs can be identified.

In another preferred embodiment, a stable hairpin secondary structure is inserted into the UTR, preferably the 5' UTR of the target gene. For example, in cases where the 5' UTR possesses IRES activity, the addition of a stable hairpin secondary structure in the 5' UTR can be used to separate cap-dependent from cap-independent translation (see, *e.g.*, Muhlrads *et al.*, 1995, *Mol. Cell. Biol.* 15(4):2145-56, the disclosure of which is incorporated by reference in its entirety). In another embodiment, an intron is inserted into a UTR (preferably, the 5' UTR) or at the 5' end of an ORF of a reporter gene. For example, but not as a limitation, in cases where an RNA possesses instability elements, an intron, *e.g.*, first intron of the human elongation factor one alpha (EF-1 alpha), can be cloned into a UTR (preferably, the 5' UTR) or a 5' end of the ORF to increase expression (see, *e.g.*, Kim *et al.*, 2002, *J Biotechnol* 93(2):183-7, the disclosure of which is incorporated by reference in its entirety). As used herein, an intron can be naturally occurring in a gene having at least two splice sites. In a preferred embodiment, an intron can be naturally occurring in a UTR. In an alternative embodiment, an intron can be naturally occurring in a heterologous gene. In an alternative embodiment, an intron can be an unnatural sequence bordered by 5' and 3' splice sites. In a preferred embodiment, both a stable hairpin secondary structure and an intron are added to the reporter gene construct. In a more preferred embodiment, the stable hairpin secondary structure is cloned into the 5' UTR and the intron is added at the 5' end of the sequence encoding the reporter polypeptide.

The reporter gene can be positioned such that the translation of that reporter gene is dependent upon the mode of translation initiation, such as, but not limited to, cap-dependent translation or cap-independent translation (*i.e.*, translation via an internal ribosome entry site). Alternatively, where the UTR contains an upstream open reading frame, the reporter gene can be positioned such that the reporter protein is translated only in the presence of a compound that shifts the reading frame of the UTR so that the formerly untranslated open

reading frame is then translated.

The reporter gene constructs can be monocistronic or multicistronic. A multicistronic reporter gene construct may encode 2, 3, 4, 5, 6, 7, 8, 9, 10 or more, or in the range of 2-5, 5-10 or 10-20 reporter genes. For example, a bicistronic reporter gene construct comprising, in the following order going downstream, a promoter, a first reporter gene, a 5' UTR of a target gene, a second reporter gene and optionally, a 3' UTR of a target gene. In such a reporter construct, the transcription of both reporter genes is capable of being driven by the promoter. In this example construct, the present invention includes the translation of the mRNA from the first reporter gene by a cap-dependent scanning mechanism and the translation of the mRNA from the second reporter gene by a cap-independent mechanism, for example by an IRES. In such a case, the IRES-dependent translation of a mRNA of the second reporter gene can be normalized against the cap-dependent translation of the first reporter gene. In a particularly preferred embodiment of the present invention, a stable hairpin secondary structure is inserted immediately downstream of the stop codon of the first reporter gene to ensure that translation of the second reporter gene cannot occur via cap-dependent translation.

Reporter genes can be expressed *in vitro* or *in vivo*. *In vivo* expression can be in a suitable bacterial or eukaryotic host. Suitable methods for expression are described by Sambrook *et al.*, *Molecular Cloning: A Laboratory Manual*, Second Edition, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y. (1989); Haymes *et al.*, *Nucleic Acid Hybridization, A Practical Approach*, IRL Press, Washington, DC (1985); or similar texts. Fusion protein or peptide molecules of the invention are preferably produced via recombinant approach. These proteins and peptide molecules can be derivatized to contain carbohydrate or other moieties (such as keyhole limpet hemocyanin, *etc.*).

#### *Linked*

As used herein, linked can mean physically linked, operably linked, flanked, or any of these in combination. In a preferred embodiment, the promoter is operably linked and physically linked to a nucleic acid sequence of the present invention.

As used herein, physically linked means that the physically linked nucleic acid sequences are located on the same nucleic acid molecule, for example a promoter can be physically linked to a reporter gene as part of a construct. If a physical linkage is proximal,

the linkage can be either direct or indirect. By way of example, a promoter that is proximally linked to a reporter gene as part of a construct can be directly linked to the reporter gene so that there is no gap between the promoter and the reporter gene. In such a case, the promoter is immediately followed by the reporter gene and there are no nucleic acid residues which do not belong to either the promoter or the reporter gene between the two elements of the construct. In an example of a promoter indirectly proximally linked to a reporter gene, nucleic acid residues which are not a part of the promoter or reporter gene exist between the promoter and reporter gene. The gap, where the nucleic acid sequence that is not derived from the promoter or reporter gene, may include for example, without limitation, a fragment, or a portion of a bovine growth hormone gene, in particular the UTR or a fragment thereof; thymidine kinase; lambda; SV40. A gap can be composed of more than approximately three stop codons. A gap can have less than five stop codons in different codon reading frames. Moreover, in one embodiment there can be multiple restriction sites, also referred to as a polylinker, between the promoter and reporter gene. In an alternative embodiment there are not multiple restriction sites, also referred to as a polylinker, between the promoter and reporter gene. In a preferred embodiment, the nucleic acid sequence in the gap is located on the nucleic acid sequence of the vector prior to cloning in the an agent of the present invention.

If the reporter gene is directly linked to a UTR of a target gene, at least one of the terminal nucleic acid residues of the reporter gene can be chemically bonded to a nucleic acid sequence from a UTR of a target gene. A UTR of a target gene (herein referred to as a "target UTR") can be the entire UTR or a fragment thereof. The reporter gene can be proximally linked indirectly to a UTR of a target gene if a terminal nucleic acid residue of the reporter gene is not chemically bonded to a nucleic acid residue from a UTR of a target gene. In a preferred embodiment, if the reporter gene is proximally linked indirectly to a UTR of a target gene, the last nucleic acid residue of the reporter gene can be about 3 residues away from a UTR of a target gene or greater than 5 but less than 20 residues away from a UTR of a target gene. If the reporter gene is directly linked to a UTR of a target gene, but that UTR of a target gene is directly followed by a UTR not in a target gene, the reporter gene is directly linked to the UTR of a target gene. In a most preferred embodiment, the reporter gene is directly linked to a UTR of a target gene as a mature mRNA, such as after a splicing event,



and can have been interrupted by a UTR not in a target gene at an earlier stage in the gene expression process.

A preferred embodiment of the present invention also provides for specific nucleic acid molecules containing a reporter gene flanked by one or more UTRs of a target gene. A UTR of a target gene refers to the nucleic acid sequence of any UTR in a target gene. In this preferred embodiment, the one or more UTRs of a target gene can be physically linked, operably linked, or operably and physically linked to the reporter gene. In a more preferred embodiment, a reporter gene is flanked by both a 5' and 3' UTR of a target gene so that compounds that effect an interaction between 5' and 3' UTRs can be identified. The effect can result in an increase or decrease in the free energy of such an interaction.

In a preferred embodiment, the reporter gene is flanked by both 5' and 3' UTRs of one or more target genes so that compounds that interfere with an interaction between the 5' and 3' UTRs can be identified. In a more preferred embodiment, the reporter gene is flanked by a 5' and 3' UTRs of one target gene, and the reporter gene is physically, operably, or physically and operably linked to the UTRs of one target gene. In a most preferred embodiment, a reporter gene is proximally linked, either directly or indirectly, to one or more UTRs of a target gene.

#### *UTRs*

Agents and constructs of the invention include nucleic acid molecules with an untranslated region (UTR). In a preferred aspect, a UTR refers to a UTR of an mRNA, *i.e.* the region of the mRNA that is not translated into protein. In a preferred embodiment, a UTR contains one or more regulatory elements that modulate untranslated region-dependent regulation of gene expression. In a particularly preferred embodiment, a UTR is a 5' UTR, *i.e.*, upstream of the coding region, or a 3' UTR, *i.e.*, downstream of the coding region. In a more preferred embodiment, a UTR contains one or more GEMs.

A UTR of the present invention can be operatively, physically, or operatively and physically linked to a target gene, target RNA, or reporter gene. In a preferred embodiment of the present invention, a UTR of the present invention is physically linked to a reporter gene. The physical, operable, or physical and operable linkage may be upstream, downstream, or internal to the reporter gene. As used herein, operably linked means that the operably linked nucleic acid sequences exhibit their deserved function. For example, a

promoter can be operably linked to a reporter gene.

In a preferred embodiment of the present invention, a UTR of the present invention is physically linked upstream of the reporter gene and another UTR is physically linked downstream of the reporter gene. In a particularly preferred embodiment, a 5' UTR of the present invention contains or consists of a GEM and is physically and operatively linked upstream of a reporter gene, and a 3' UTR is physically and operatively linked downstream of the reporter gene. In an alternatively preferred embodiment, a 3' UTR of the present invention contains or consists of a GEM and is physically and operatively linked downstream of a reporter gene, and a 5' UTR is physically and operatively linked upstream of the reporter gene. In an alternatively preferred embodiment, a 5' UTR of the present invention contains or consists of a GEM and is physically and operatively linked upstream of a reporter gene, and a 3' UTR of the present invention contains or consists of a GEM and is physically and operatively linked downstream of the reporter gene. One or more GEMs in a 5' UTR, in a 3' UTR, or both in the 5' and 3' UTRs can act independently or dependently of linked nucleic acid sequence.

In a preferred embodiment of the present invention, a UTR of the present invention is physically linked to reporter gene containing an intron. In a more preferred embodiment of the present invention, a UTR of the present invention containing a GEM is physically linked to a reporter gene containing an intron. In a preferred embodiment of the present invention, a 5' UTR of the present invention is physically linked upstream of a reporter gene and contains an intron internal to the UTR. In a preferred embodiment of the present invention, a UTR of the present invention is physically linked upstream of a reporter gene and a UTR is physically linked downstream of the reporter gene.

A gene can include regions preceding and following a nucleic acid sequence encoding a polypeptide as well as introns between the exons of the coding region. A typical mRNA contains a 5' cap, a 5' untranslated region ("5' UTR") upstream of a start codon, an open reading frame, which is also referred to as a coding sequence that encodes a stable RNA or a functional protein, a 3' untranslated region ("3' UTR") downstream of the termination codon, and a poly(A) tail. A nucleic acid of the present invention can include a UTR containing a GEM, a GEM, a fragment of either, or a complement of any of these. In a preferred

embodiment, a *cis*-dependent RNA-based GEM maps to the 5' UTR, the 3' UTR, or the 5' UTR and 3' UTR.

### *GEMs*

As referred to herein, a GEM is a gene expression modulator that regulates expression of a target gene after transcription. In one aspect, a GEM is not a full-length sequence of a UTR from a target gene (hereafter referred to as "a target UTR"). In a preferred aspect, a GEM is not a full-length 5' UTR or a full-length 3' UTR. A GEM can include the nucleic acid sequence involved in modulation of expression as a result of interaction between UTRs, preferably the interaction between a 5' UTR and a 3' UTR from the same gene, a UTR pair. In one embodiment, a GEM in one target gene can have primary nucleic acid sequence similarity to a GEM in a different target gene. Alternatively, there may not be any primary nucleic acid sequence similarity in GEMs of similar function. In a preferred embodiment, a GEM in one target gene can have a secondary, tertiary, or secondary and tertiary structure similar to a GEM in a different target gene. Examples of GEMs include, but are not limited to, IRES elements, upstream ORFs, and AREs.

In one embodiment, a GEM of the present invention is a nucleic acid sequence in a UTR, which modulates UTR-dependent gene expression after transcription of the gene. A GEM can be a nucleic acid sequence located anywhere in a target gene. Examples of 5' UTR regulatory elements, such as GEMs of the present invention, include the iron response element ("IRE"), internal ribosome entry site ("IRES"), upstream open reading frame ("uORF"), male specific lethal element ("MSL-2"), G-quartet element, and 5'-terminal oligopyrimidine tract ("TOP") (reviewed in Keene & Tenenbaum, 2002, Mol Cell 9:1161 and Translational Control of Gene Expression, Sonenberg, Hershey, and Mathews, eds., 2000, CSHL Press). Examples of 3' UTR regulatory elements, such as GEMs of the present invention, include AU-rich elements ("AREs"), Selenocysteine insertion sequence ("SECIS"), histone stem loop, cytoplasmic polyadenylation elements ("CPEs"), nanos translational control element, amyloid precursor protein element ("APP"), translational regulation element ("TGE"), direct repeat element ("DRE"), bruno element ("BRE"), 15-lipoxygenase differentiation control element (15-LOX-DICE), and G-quartet element (reviewed in Keene & Tenenbaum, 2002, Mol Cell 9:1161). GEMs include nucleic acid sequences in a UTR that modulate other GEM sequences.

By way of example, a GEM in the 5' UTR of a target gene can modulate the GEM-dependent expression of a GEM in the same or another UTR, for example, a GEM in the 3' UTR of the same target gene. In a particularly preferred embodiment, a GEM can consist of the interaction between sequences of the 5' and 3' UTR of the same target gene where the GEM activity requires the presence of both the 5' and 3' UTR whose sequence elements cannot function independently). GEMs of the present invention can be located in any position within a construct and not limited to the 5' UTR or 3' UTR regions of a molecule. A GEM of the present invention can be operatively, physically, or operatively and physically linked to a UTR. In an alternative embodiment of the present invention, a GEM of the present invention is a UTR of the present invention.

In one embodiment of the present invention, a GEM is located between about 1 to about 100 residues upstream from the initiation codon of an open reading frame in a mRNA, between about 150 to about 250 residues upstream from the initiation codon, or between about 300 to about 500 residues upstream from the initiation codon. In a most preferred embodiment, a GEM is within about 30 residues upstream from the initiation codon. In addition to or independent of other GEMs in a nucleic acid molecule, a GEM of the present invention can be located between about 1 to about 100 residues downstream from the stop codon of an open reading frame in a mRNA, between about 150 to about 250 residues downstream from the stop codon, or between about 300 to about 500 residues downstream from the stop codon. In a preferred embodiment, a GEM is within about 30 residues downstream from the stop codon.

Further examples of embodiments of the present invention include a GEM within about 1000 residues upstream from the 5' end of a main ORF, within about 500 residues upstream from the 5' end of a main ORF, or within about 200 residues upstream from the 5' end of a main ORF, or within about 100 residues upstream from the 5' end of a main ORF. A GEM of the present invention can also be located within about 1000 residues downstream from the 3' end of a main ORF, within about 500 residues downstream from the 3' end of a main ORF, or within about 200 residues downstream from the 3' end of a main ORF or within about 100 residues downstream from the 3' end of a main ORF. In a preferred embodiment, a GEM is about 5 residues down stream from the stop codon of a main ORF.

Constructs of the present invention can have more or fewer components than described above. For example, constructs of the present invention can include genetic elements, including but not limited to, 3' transcriptional terminators, 3' polyadenylation signals, other untranslated nucleic acid sequences, transit or targeting sequences, selectable or screenable markers, promoters, enhancers, and operators, as desired. Constructs of the present invention can also contain a promoterless gene that may utilize an endogenous promoter upon insertion into a host cell chromosome.

Alternatively, sequences encoding nucleic acid molecules of the present invention can be cloned into a vector for the production of an mRNA probe. Such vectors are known in the art, are commercially available, and can be used to synthesize RNA probes in vitro by addition of labeled nucleotides and an appropriate RNA polymerase such as T7, T3, or SP6. These procedures can be conducted using a variety of commercially available kits (for example, Amersham Biosciences Inc., Piscataway, NJ; and Promega Co, Madison, WI).

#### Modulation of Gene Expression by Nucleic Acid Molecules of the Present Invention

Modulation of gene expression can result in more or less gene expression. Many approaches for modulating gene expression using nucleic acid molecules of the present invention are known to one skilled in the art. For example, over-expression of a gene product can be the result from transfection of a construct of the present invention into a mammalian cell. Similarly, down-regulation can be the result from transfection of a construct of the present invention into a mammalian cell. Other non-limiting examples include anti-sense techniques like RNA interference (RNAi), transgenic animals, hybrids, and ribozymes. The following examples are provided by way of illustration, and are not intended to be limiting of the present invention.

##### *Cellular Mechanisms*

As used herein, the term "UTR-dependent expression" refers to the regulation of gene expression through a UTR at the level of mRNA expression, *i.e.*, after transcription of the gene has begun until the protein or the RNA product(s) encoded by the gene has been degraded. In a preferred embodiment, the term "UTR-dependent expression" refers to the regulation of mRNA stability or translation. In a more preferred embodiment, the term "UTR-dependent expression" refers to the regulation of gene expression through regulatory

elements present in a UTR. Altering the sequence of a GEM within a UTR of target gene can change the amount of UTR-dependent expression observed for that target gene.

As used herein, a “UTR-affected mechanism” is a cellular mechanism that discriminates between UTRs based on their nucleic acid sequence or based on properties that are a function of their sequence such as the secondary, tertiary, or quaternary structure. In an embodiment of the present invention, a UTR-affected mechanism discriminates between UTRs based on a UTR sequence-dependent higher order complex assembly of trans-acting factors. Modulation of the UTR-dependent expression of a target gene can be due to a change in how a UTR-affected mechanism acts on the target gene. For example, a UTR in a target gene can contain an IRES, which affects target gene expression via a UTR-affected mechanism.

In a preferred embodiment, a UTR-affected mechanism can be a main ORF-independent mechanism. As used herein, a “main ORF-independent mechanism” refers to a cellular pathway or process, wherein at least one step relates to gene expression and is not dependent on the nucleic acid sequence of the main open reading frame. In a preferred embodiment, a UTR-affected mechanism is a main ORF-independent, UTR-affected mechanism.

In order to exclude the possibility that a particular compound is functioning solely by modulating the expression of a target gene in a UTR-independent manner, one or more mutations may be introduced into the UTRs operably linked to a reporter gene and the effect on the expression of the reporter gene in a reporter gene-based assay described herein can be determined. For example, a reporter gene construct comprising the 5' UTR of a target gene may be mutated by deleting a fragment of the 5' UTR of the target gene or substituting a fragment of the 5' UTR of the target gene with a fragment of the 5' UTR of another gene and measuring the expression of the reporter gene in the presence and absence of a compound that has been identified in screening assays of the present invention or of an assay well known to the skilled artisan. If the deletion of a fragment of the 5' UTR of the target gene or the substitution of a fragment of the 5' UTR of the target gene with a fragment of the 5' UTR of another gene affects the ability of the compound to modulate the expression of the reporter gene, then the fragment of the 5' UTR deleted or substituted plays a role in the ability of the compound to regulate reporter gene expression and the regulation, at least in part, is UTR-

dependent.

Alternatively or in conjunction with the tests described above, the possibility that a particular compound is functioning solely by modulating the expression of a target gene in an UTR-independent manner can be determined by changing the vector utilized as a reporter construct. The UTRs flanked by a reporter gene from the first reporter construct in which an effect on reporter gene expression was detected following exposure to a compound may be inserted into a new reporter construct that has, *e.g.*, different transcriptional regulation elements (*e.g.*, a different promoter) and a different selectable marker. The level of reporter gene expression in the presence of the compound can be compared to the level of reporter gene expression in the absence of the compound or in the presence of a control (*e.g.*, PBS). If there is no change in the level of expression of the reporter gene in the presence of the compound relative to the absence of the compound or in the presence of a control, then the compound probably is functioning in an UTR-independent manner.

By way of further example, additional tests can be used to evaluate that a particular compound functions by modulating the expression of a target gene in an UTR-independent manner. This can be done, for example, by measuring the effect of the compound when the reporter gene is operably linked to UTRs from another target gene. The potency with which the compound effects the level of reporter gene expression operably linked to the original UTRs can be compared to the potency with which the compound effects the level of reporter gene expression operably linked to the control UTRs. If the compound is active only when the original UTRs are operably linked to the reporter gene and shows a significant decrease in activity when the control UTRs are operably linked to the reporter gene, then the compound is a candidate compound that functions in a UTR-independent manner.

Compounds, identified in assays of the present invention, that are capable of modulating UTR-dependent expression of a target gene (for convenience referred to herein as a "lead" compound) can be further tested for UTR-dependent binding to the target RNA (which contains at least one UTR, and preferably at least one element of an UTR, for example a GEM). Furthermore, by assessing the effect of a compound on target gene expression, *cis*-acting elements, *i.e.*, specific nucleotide sequences, that are involved in UTR-dependent expression may be identified. RNA binding assays, subtraction assays, and

expressed protein concentration and activity assays are examples methods to determine UTR-dependent expression of a gene.

### *Hybrids*

In one aspect of the present invention, a hybrid of a compound and a GEM of the present invention is a hybrid formed between two non-identical molecules. In a preferred aspect, a hybrid can be formed between two nucleic acid molecules. For example, a hybrid can be formed between two ribonucleic acid molecules, between a ribonucleic acid molecule and a deoxyribonucleic acid molecule, or between derivatives of either. In alternative embodiment, a hybrid can be formed between a nucleic acid of the present invention and a non-nucleic acid molecule. In a preferred embodiment, a hybrid can be formed between a nucleic acid molecule and a non-nucleic acid molecule, for example, a polypeptide or a non-peptide therapeutic agent.

### *Ribozymes*

In one aspect of the present invention, the activity or expression of a gene is regulated by designing trans-cleaving catalytic RNAs (ribozymes) specifically directed to a nucleic acid molecule of the present invention. In an alternate aspect, the activity or expression of a gene is regulated by designing trans-cleaving catalytic RNAs (ribozymes) specifically directed to a nucleic acid molecule of the present invention.

Ribozymes are RNA molecules possessing endoribonuclease activity. Ribozymes are specifically designed for a particular target, and the target message contains a specific nucleotide sequence. They are engineered to cleave any RNA species site-specifically in the background of cellular RNA. The cleavage event renders the mRNA unstable and prevents protein expression. Importantly, ribozymes can be used to inhibit expression of a gene of unknown function for the purpose of determining its function in an *in vitro* or *in vivo* context, by detecting a phenotypic effect.

One commonly used ribozyme motif is the hammerhead, for which the substrate sequence requirements are minimal. Design of the hammerhead ribozyme, and the therapeutic uses of ribozymes, are disclosed in Usman *et al.*, *Current Opin. Strict. Biol.* 6:527-533 (1996). Ribozymes can also be prepared and used as described in Long *et al.*, *FASEB J.* 7:25 (1993); Symons, *Ann. Rev. Biochem.* 61:641 (1992); Perrotta *et al.*, *Biochem.*



31:16-17 (1992); Ojwang *et al.*, *PNAS* 89:10802-10806 (1992); and U.S. Patent No. 5,254,678.

Ribozyme cleavage of HIV-I RNA, methods of cleaving RNA using ribozymes, methods for increasing the specificity of ribozymes, and the preparation and use of ribozyme fragments in a hammerhead structure are described in U.S. Patent Nos. 5,144,019; 5,116,742; and 5,225,337 and Koizumi *et al.*, *Nucleic Acid Res.* 17:7059-7071 (1989). Preparation and use of ribozyme fragments in a hairpin structure are described by Chowrira and Burke, *Nucleic Acids Res.* 20:2835 (1992). Ribozymes can also be made by rolling transcription as described in Daubendiek and Kool, *Nat. Biotechnol.* 15(3):273-277 (1997).

The hybridizing region of the ribozyme may be modified or may be prepared as a branched structure as described in Horn and Urdea, *Nucleic Acids Res.* 17:6959-67 (1989). The basic structure of the ribozymes may also be chemically altered in ways familiar to those skilled in the art, and chemically synthesized ribozymes can be administered as synthetic oligonucleotide derivatives modified by monomeric units. In a therapeutic context, liposome mediated delivery of ribozymes improves cellular uptake, as described in Birikh *et al.*, *Eur. J. Biochem.* 245:1-16 (1997).

Ribozymes of the present invention also include RNA endoribonucleases (hereinafter "Cech-type ribozymes") such as the one which occurs naturally in *Tetrahymena thermophila* (known as the IVS, or L-19 IVS RNA) and which has been extensively described by Thomas Cech and collaborators (Zaug *et al.*, *Science* 224:574-578 (1984); Zaug and Cech, *Science* 231:470-475 (1986); Zaug *et al.*, *Nature*, 324:429-433 (1986); WO 88/04300; Been and Cech, *Cell* 47:207-216 (1986)). The Cech-type ribozymes have an eight base pair active site which hybridizes to a target RNA sequence whereafter cleavage of the target RNA takes place. The invention encompasses those Cech-type ribozymes which target eight base-pair active site sequences that are present in a target gene.

Ribozymes can be composed of modified oligonucleotides (*e.g.*, for improved stability, targeting, etc.) and should be delivered to cells which express the target gene *in vivo*. A preferred method of delivery involves using a DNA construct "encoding" the ribozyme under the control of a strong constitutive pol III or pol II promoter, so that transfected cells will produce sufficient quantities of the ribozyme to destroy endogenous

messages and inhibit translation. Because ribozymes, unlike antisense molecules, are catalytic, a lower intracellular concentration is required for efficiency.

Using the nucleic acid sequences of the invention and methods known in the art, ribozymes are designed to specifically bind and cut the corresponding mRNA species. Ribozymes thus provide a method to inhibit the expression of any of the proteins encoded by the disclosed nucleic acids or their full-length genes. The nucleic acid sequence of the full-length gene need not be known in order to design and use specific inhibitory ribozymes. In the case of a nucleic acid or cDNA of unknown function, ribozymes corresponding to the specific nucleotide sequence can be tested *in vitro* for efficacy in cleaving the target transcript. Those ribozymes that effect cleavage *in vitro* are further tested *in vivo*. The ribozyme can also be used to generate an animal model for a disease, as described in Birikh *et al.*, *Eur. J. Biochem.* 245:1-16 (1997). An effective ribozyme is used to determine the function of the gene of interest by blocking its expression and detecting a phenotypic change in the cell. Where the gene is found to be a mediator in a disease, an effective ribozyme is designed and delivered in a gene therapy for blocking expression of the gene.

Therapeutic and functional genomic applications of ribozymes begin with knowledge of a portion of the coding sequence of the gene to be inhibited. Thus, for many genes, a partial nucleic acid sequence provides adequate sequence for constructing an effective ribozyme. A target cleavage site is selected in the target sequence, and a ribozyme is constructed based on the 5' and 3' nucleotide sequences that flank the cleavage site. Retroviral vectors are engineered to express monomeric and multimeric hammerhead ribozymes targeting the mRNA of the target coding sequence. These monomeric and multimeric ribozymes are tested *in vitro* for an ability to cleave the target mRNA. A cell line is stably transduced with the retroviral vectors expressing the ribozymes, and the transduction is confirmed by Northern blot analysis and reverse-transcription polymerase chain reaction (RT-PCR). The cells are screened for inactivation of the target mRNA by such indicators as reduction of expression of disease markers or reduction of the gene product of the target mRNA.

#### Cells and Organisms

Nucleic acid molecules that may be used in cell transformation or transfection can be any of the nucleic acid molecules of the present invention. Nucleic acid molecules of the

present invention can be introduced into a cell or organism. A heterologous nucleic acid molecule can be an RNA molecule produced in a different cell or produced by *in vitro* transcription (Ambion, Inc., Austin, TX) and transfected directly into a cell of interest.

A host cell strain can be chosen for its ability to modulate the expression of the inserted sequences, to process an expressed reporter gene in the desired fashion, or based on the expression levels of endogenous or heterologous target genes. Mammalian cell lines available as hosts for expression are known in the art and include many immortalized cell lines available from the American Type Culture Collection (ATCC, Manassas, VA), such as HeLa cells, Chinese hamster ovary (CHO) cells, baby hamster kidney (BHK) cells and a number of other cell lines. Non-limiting examples of suitable mammalian host cell lines include those shown below in Table 1.

Table 1: Mammalian Host Cell Lines

Host Cell	Origin	Source
HepG-2	Human Liver Hepatoblastoma	ATCC HB 8065
CV-1	African Green Monkey Kidney	ATCC CCL 70
LLC-MK <sub>2</sub>	Rhesus Monkey Kidney	ATCC CCL 7
3T3	Mouse Embryo Fibroblasts	ATCC CCL 92
AV12-664	Syrian Hamster	ATCC CRL 9595
HeLa	Human Cervix Epitheloid	ATCC CCL 2
RPMI8226	Human Myeloma	ATCC CCL 155
H4IIEC3	Rat Hepatoma	ATCC CCL 1600
C127I	Mouse Fibroblast	ATCC CCL 1616
293	Human Embryonal Kidney	ATCC CRL 1573
HS-Sultan	Human Plasma Cell Plasmocytoma	ATCC CCL 1484
BHK-21	Baby Hamster Kidney	ATCC CCL 10
CHO-K1	Chinese Hamster Ovary	ATCC CCL 61

In a preferred aspect, cells of the present invention can be cells of an organism. In a more preferred aspect, the organism is a mammal. In a most preferred aspect, the mammal is a human. In another more preferred aspect, the organism is a non-human mammal,

preferably a mouse, rat, or a chimpanzee. In one aspect of the present invention, cells can be pluripotent or differentiated.

A nucleic acid of the present invention can be naturally occurring in the cell or can be introduced using techniques such as those described in the art. There are many methods for introducing transforming DNA segments into cells, but not all are suitable for delivering DNA to eukaryotic cells. Suitable methods include any method by which DNA can be introduced into a cell, such as by direct delivery of DNA, by desiccation/inhibition-mediated DNA uptake, by electroporation, by agitation with silicon carbide fibers, by acceleration of DNA coated particles, by chemical transfection, by lipofection or liposome-mediated transfection, by calcium chloride-mediated DNA uptake, etc. For example, without limitation, Lipofectamine<sup>®</sup> (Invitrogen Co., Carlsbad, CA) and Fugene<sup>®</sup> (Hoffmann-La Roche Inc., Nutley, NJ) can be used for transfection of nucleic acid molecules, such as constructs and small interfering RNAs (siRNA), into several mammalian cells. Alternatively, in certain embodiments, acceleration methods are preferred and include, for example, microprojectile bombardment and the like. Within the scope of this invention, the transfected nucleic acids of the present invention may be expressed transiently or stably. Such transfected cells can be in a two- or three-dimensional cell culture system or in an organism.

For example, without limitation, the construct may be an autonomously replicating construct, *i.e.*, a construct that exists as an extrachromosomal entity, the replication of which is independent of chromosomal replication, *e.g.*, a plasmid, an extrachromosomal element, a minichromosome, or an artificial chromosome. The construct may contain any approach for assuring self-replication. For autonomous replication, the construct may further comprise an origin of replication enabling the construct to replicate autonomously in the host cell. Alternatively, the construct may be one which, when introduced into the cell, is integrated into the genome and replicated together with the chromosome(s) into which it has been integrated. This integration may be the result of homologous or non-homologous recombination.

Integration of a construct or nucleic acid into the genome by homologous recombination, regardless of the host being considered, relies on the nucleic acid sequence of the construct. Typically, the construct contains nucleic acid sequences for directing integration by homologous recombination into the genome of the host. These nucleic acid

sequences enable the construct to be integrated into the host cell genome at a precise location or locations in one or more chromosomes. To increase the likelihood of integration at a precise location, there should be preferably two nucleic acid sequences that individually contain a sufficient number of nucleic acids, preferably 400 residues to 1500 residues, more preferably 800 residues to 1000 residues, which are highly homologous with the corresponding host cell target sequence. This enhances the probability of homologous recombination. These nucleic acid sequences may be any sequence that is homologous with a host cell target sequence and, furthermore, may or may not encode proteins.

Stable expression is preferred for long-term, high-yield production of recombinant proteins. For example, to generate cell lines that stably express a reporter gene, cell lines can be transformed using expression constructs that can contain viral origins of replication and/or endogenous expression elements and a selectable marker gene on the same or on a separate construct. Following the introduction of the construct, cells can be allowed to grow for 1-2 days in an enriched medium before they are switched to a selective medium. The purpose of the selectable marker is to confer resistance to selection, and its presence allows growth and recovery of cells that successfully express the introduced construct. Resistant clones of stably transformed cells can be proliferated using tissue culture techniques appropriate to the cell type. See, for example, *Animal Cell Culture*, R.I. Freshney, ed., 1986.

Any number of selection systems can be used to recover transformed cell lines. These include, but are not limited to, the herpes simplex virus thymidine kinase (Wigler *et al.*, *Cell*11:223-32 (1977)) and adenine phosphoribosyltransferase (Lowy *et al.*, *Cell*22:817-23 (1980)) genes which can be employed in *tk* or *aprt* cells, respectively. Also, antimetabolite, antibiotic, or herbicide resistance can be used as the basis for selection. For example, *dhfr* confers resistance to methotrexate (Wigler *et al.*, *Proc. Natl. Acad. Sci.*77:3567-70 (1980)), *npt* confers resistance to the aminoglycosides, neomycin and G-418 (Colbere-Garapin *et al.*, *J. Mol. Biol.*150: 1-14 (1981), and *als* and *pat* confer resistance to chlorsulfuron and phosphinotricin acetyltransferase, respectively. Additional selectable genes have been described. For example, *trpB* allows cells to utilize indole in place of tryptophan, and *hisD* allows cells to utilize histinol in place of histidine (Hartman & Mulligan, *Proc. Natl. Acad. Sci.*85:8047-51 (1988)). Visible markers such as anthocyanins,  $\beta$ -glucuronidase and its

substrate GUS, and luciferase and its substrate luciferin, can be used to identify transformants and to quantify the amount of transient or stable protein expression attributable to a specific construct system (Rhodes *et al.*, *Methods Mol. Biol.* 55:121-131 (1995)).

Although the presence of marker gene expression suggests that a reporter gene is also present, its presence and expression may need to be confirmed. For example, if a sequence encoding a reporter gene is inserted within a marker gene sequence, transformed cells containing sequences that encode a reporter gene can be identified by the absence of marker gene function. Alternatively, a marker gene can be placed in tandem with a sequence encoding a reporter gene under the control of a single promoter. Expression of the marker gene in response to induction or selection usually indicates expression of a reporter gene.

Alternatively, host cells which contain and express a reporter gene and can be identified by a variety of procedures known to those of skill in the art. These procedures include, but are not limited to, DNA-DNA or DNA-RNA hybridizations and protein bioassay or immunoassay techniques that include membrane, solution, or chip-based technologies for the detection and/or quantification of nucleic acid or protein. For example, the presence of a reporter gene can be detected by DNA-DNA or DNA-RNA hybridization or amplification using probes or fragments or fragments of polynucleotides encoding a reporter gene. Nucleic acid amplification-based assays involve the use of oligonucleotides selected from sequences encoding a reporter gene to detect transformants that contain a reporter gene.

#### Screening Methods of the Present Invention

Another aspect of the present invention includes screening methods to identify agents and compounds that modulate gene expression and can result in more or less gene expression. Many methods for screening agents and compounds that modulating gene expression are known to one skilled in the art. For example, over-expression of a gene product can be the result from transfection of a construct of the present invention into a mammalian cell. Similarly, down-regulation can be the result from transfection of a construct of the present invention into a mammalian cell. Other non-limiting examples include anti-sense techniques like RNA interference (RNAi), transgenic animals, hybrids, and ribozymes. The following examples are provided by way of illustration, and are not intended to be limiting of the present invention.

### *Compound*

The present invention includes methods for screening compounds capable of modulating gene expression.

Any compound can be screened in an assay of the present invention. In an embodiment, a compound includes a nucleic acid or a non-nucleic acid, such as a polypeptide or a non-peptide therapeutic agent. In a preferred embodiment, a nucleic acid can be a polynucleotide, a polynucleotide analog, a nucleotide, or a nucleotide analog. In a more preferred embodiment, a compound can be an antisense oligonucleotide, which are nucleotide sequences complementary to a specific DNA or RNA sequence of the present invention. Preferably, an antisense oligonucleotide is at least 11 nucleotides in length, but can be at least 12, 15, 20, 25, 30, 35, 40, 45, or 50 or more nucleotides long. Longer sequences also can be used. Antisense oligonucleotides can be deoxyribonucleotides, ribonucleotides, or a combination of both.

Nucleic acid molecules, including antisense oligonucleotide molecules, can be provided in a DNA construct and introduced into a cell. Nucleic acid molecules can be antisense or sense and double- or single-stranded. In a preferred embodiment, nucleic acid molecules can be interfering RNA (RNAi) or microRNA (miRNA). In a preferred embodiment, the dsRNA is 20-25 residues in length, termed small interfering RNAs (siRNA).

Oligonucleotides can be synthesized manually or by an automated synthesizer, by covalently linking the 5' end of one nucleotide with the 3' end of another nucleotide with non-phosphodiester internucleotide linkages such as alkylphosphonates, phosphorothioates, phosphorodithioates, alkylphosphonothioates, alkylphosphonates, phosphoramidates, phosphate esters, carbamates, acetamidate, carboxymethyl esters, carbonates, and phosphate triesters. See Brown, 1994 *Meth. Mol. Biol.* vol. 20:1-8; Sonveaux, 1994. *Meth. Mol. Biol.* Vol. 26:1-72; and Uhlmann *et al.*, 1990. *Chem. Rev.* vol. 90:543-583. Salts, esters, and other pharmaceutically acceptable forms of such compounds are also encompassed.

In a preferred embodiment, a compound can be a peptide, polypeptide, polypeptide analog, amino acid, or amino acid analog. Such a compound can be synthesized manually or by an automated synthesizer. Any peptide, polypeptide, polypeptide analog, amino acid, or amino acid analog can be involved in UTR-dependent modulation of gene expression

mediated by a GEM. Compounds detected by an assay of the present invention can modulate interactions of a GEM including of a UTR-complex containing a protein or a ribonucleoprotein. Such a compound can increase or decrease the interaction of a GEM and protein or protein complex.

A compound can be a member of a library of compounds. In a specific embodiment, the compound is selected from a combinatorial library of compounds comprising peptoids; random biooligomers; diversomers such as hydantoins, benzodiazepines and dipeptides; vinylogous polypeptides; nonpeptidal peptidomimetics; oligocarbamates; peptidyl phosphonates; peptide nucleic acid libraries; antibody libraries; carbohydrate libraries; and small organic molecule libraries. In a preferred embodiment, the small organic molecule libraries are libraries of benzodiazepines, isoprenoids, thiazolidinones, metathiazanones, pyrrolidines, morpholino compounds, or diazepindiones.

In another embodiment, a compound can have a molecular weight less than about 10,000 grams per mole, less than about 5,000 grams per mole, less than about 1,000 grams per mole, less than about 500 grams per mole, less than about 100 grams per mole, and salts, esters, and other pharmaceutically acceptable forms of such compounds.

Compounds can be evaluated comprehensively for cytotoxicity. The cytotoxic effects of the compounds can be studied using cell lines, including for example 293T (kidney), HuH7 (liver), and Hela cells over about 4, 10, 16, 24, 36 or 72-hour periods. In addition, a number of primary cells such as normal fibroblasts and peripheral blood mononuclear cells (PBMCs) can be grown in the presence of compounds at various concentrations for about 4 days. Fresh compound can be added every other day to maintain a constant level of exposure with time. The effect of each compound on cell-proliferation can be determined by CellTiter 96® AQueous One Solution Cell Proliferation Assay (Promega Co, Madison, WI) and [<sup>3</sup>H]-thymidine incorporation. Treatment of some cells with some of the compounds may have cytostatic effects. A selective index (ratios of CC<sub>50</sub> in cytotoxicity assays to the EC<sub>50</sub> in ELISA or FACS or the reporter gene assays) for each compound can be calculated for all of the UTR-reporters and protein inhibition assays. Compounds exhibiting substantial selective indices can be of interest and can be analyzed further in the functional assays.

The structure of a compound can be determined by any well-known method such as mass spectroscopy, NMR, vibrational spectroscopy, or X-ray crystallography as part of a



method of the present invention.

Compounds can be pharmacologic agents already known in the art or can be compounds previously unknown to have any pharmacological activity. The compounds can be naturally occurring or designed in the laboratory. They can be isolated from microorganisms, animals, or plants, and can be produced recombinantly, or synthesized by chemical methods known in the art. If desired, compounds can be obtained using any of the numerous combinatorial library methods known in the art, including but not limited to, biological libraries, spatially addressable parallel solid phase or solution phase libraries, synthetic library methods requiring deconvolution, the "one-bead one-compound" library method, and synthetic library methods using affinity chromatography selection. Methods for the synthesis of molecular libraries are well known in the art (see, for example, DeWitt et al., Proc. Natl. Acad. Sci. U.S.A. 90, 6909, 1993; Erb et al. Proc. Natl. Acad. Sci. U.S.A. 91, 11422, 1994; Zuckermann et al., J. Med. Chem. 37, 2678, 1994; Cho et al., Science 261, 1303, 1993; Carell et al., Angew. Chem. Int. Ed. Engl. 33, 2059, 1994; Carell et al., Angew. Chem. Int. Ed. Engl. 33, 2061; Gallop et al., J. Med. Chem. 37, 1233, 1994). Libraries of compounds can be presented in solution (see, e.g., Houghten, BioTechniques 13, 412-421, 1992), or on beads (Lam, Nature 354, 82-84, 1991), chips (Fodor, Nature 364, 555-556, 1993), bacteria or spores (Ladner, U.S. Pat. No. 5,223,409), plasmids (Cull et al., Proc. Natl. Acad. Sci. U.S.A. 89, 1865-1869, 1992), or phage (Scott & Smith, Science 249, 386-390, 1990; Devlin, Science 249, 404-406, 1990); Cwirla et al., Proc. Natl. Acad. Sci. 97, 6378-6382, 1990; Felici, J. Mol. Biol. 222, 301-310, 1991; and Ladner, U.S. Pat. No. 5,223,409).

Methods of the present invention for screening compounds can select for compounds capable of modulating gene expression, which are capable of directly binding to a ribonucleic acid molecule transcribed from a target gene. In a preferred embodiment, a compound identified in accordance with the methods of the present invention may be capable of binding to one or more *trans*-acting factors (such as, but not limited to, proteins) that modulate UTR-dependent expression of a target gene. In another preferred embodiment, a compound identified in accordance with the methods of invention may disrupt an interaction between the 5' UTR and the 3' UTR.

Compounds can be tested using *in vitro* assays (e.g., cell-free assays) or *in vivo* assays (e.g., cell-based assays) well known to one of skill in the art or as provided in the present

invention. A compound that modulates expression of a target gene can be determined from the methods provided in the present invention. A UTR of the present invention includes UTRs capable of modulating gene expression in the presence, in the absence, or in the presence and absence of a compound. In a preferred embodiment, the effect of a compound on the expression of one or more genes can be determined utilizing assays well known to one of skill in the art or provided by the present invention to assess the specificity of a particular compound's effect on the UTR-dependent expression of a target gene. In a more preferred embodiment, a compound has specificity for a plurality of genes. In another more preferred embodiment, a compound identified utilizing the methods of the present invention is capable of specifically effect the expression of only one gene or, alternatively, a group of genes within the same signaling pathway. Compounds identified in the assays of the present invention can be tested for biological activity using host cells containing or engineered to contain the target RNA element involved in UTR-dependent gene expression coupled to a functional readout system.

#### *Screening assays*

The present invention includes and provides for assays capable of screening for compounds capable of modulating gene expression. In a preferred aspect of the present invention, an assay is an *in vitro* assay. In another aspect of the present invention, an assay is an *in vivo* assay. In another preferred aspect of the present invention, an assay measures translation. In a preferred aspect of the present invention, the assay includes a nucleic acid molecule of the present invention or a construct of the present invention. A nucleic acid molecule or construct of the present invention includes, without limitation, a GEM, or a sequence that differs from any of the residues in a GEM in that the nucleic acid sequence has been deleted, substituted, or added in a manner that does not alter the function. The present invention also provides fragments and complements of all the nucleic acid molecules of the present invention.

In one aspect of a preferred, present invention, the activity or expression of a reporter gene is modulated. Modulated means increased or decreased expression during any point before, after, or during translation. In a preferred embodiment, activity or expression of a reporter gene is modulated during translation. For example, inhibition of translation of the reporter gene can modulate expression. In an alternative example, the expression level of a

reporter gene is modulated if the steady-state level of the expressed protein decreased even though translation was not inhibited. As a further example, a change in the half-life of a mRNA can modulate expression.

In an alternative embodiment, modulated activity or expression of a reporter gene means increased or decreased expression during any point before, during, or after translation.

In a more preferred aspect, the activity or expression of a reporter gene or a target gene is modulated by greater than 30%, 40%, 50%, 60%, 70%, 80% or 90% in the presence of a compound. In a highly preferred aspect, more of an effect is observed in cancer cells.

Expression of a reporter gene can be detected with, for example, techniques known in the art. Translation of a reporter gene can be detected *in vitro* or *in vivo*. In detection assays, either the compound or the reporter gene can comprise a detectable label, such as a fluorescent, radioisotopic or chemiluminescent label or an enzymatic label, such as horseradish peroxidase, alkaline phosphatase, or luciferase.

Using an assay of the present invention, a compound that affects a UTR or multiple UTRs from one target gene can be determined. In a preferred embodiment, a compound that affects the 5' UTR, 3' UTR, or the 5' and 3' UTRs from a single target gene can be detected. In another preferred embodiment, the 5' and 3' UTRs from multiple target genes are each reacted with multiple compounds, and an effect of a compound on a UTR can be detected.

In an assay of the present invention, the result of one or more UTRs being affected by a compound is qualitatively, quantitatively, or qualitatively and quantitatively determined based on the modulation of expression from a reporter gene operatively linked to the UTRs. The modulation of expression from a reporter gene operatively linked to the UTR can be relative to the expression from a reporter gene operatively linked to the UTR in the absence of the compound, in comparison to a different dosage of the same compound, in comparison to another compound, in comparison to the reaction of another UTR/compound effect, or by combining the results of these comparisons.

A compound can be reacted with one or multiple UTRs operatively linked to a reporter gene. If the compound modulates the expression of a reporter gene operatively linked to a UTR, the compound can be determined to be specifically active, nonspecifically active, or inactive with respect to the one or more UTRs being tested. The compound is specifically active if it modulates the expression of a reporter gene operatively linked to some

UTRs, but not all UTRs, being tested. The compound is nonspecifically active if it similarly modulates the expression of a reporter gene operatively linked to all of the UTRs being tested. Whether the compound similarly modulates the expression of a reporter gene operatively linked to more than one UTR can be determined statistically. Similar modulation occurs when the effect of the compound modulates the reporter gene expression within an order of magnitude for the UTRs tested. The compound is inactive if it does not modulate the expression from a reporter gene operatively linked to any of the UTRs tested.

One or more UTRs can be tested with one or more compounds. In a preferred embodiment, there can be any number of UTRs tested, for example without limitation, one, ten, hundreds, thousands, tens of thousands, or hundreds of thousands of UTRs or UTR pairs, where UTR pairs refers to a 5' UTR and a 3' UTR from the same target gene. In a preferred embodiment, a single pair of UTRs is reacted with about 2,000 – about 5,000 compounds. In a more preferred embodiment, each UTR reacts with each compound at about 3 – about 7 concentrations, for example, without limitation, using a 4-point 10-fold dose-response.

Compounds of the present invention can be categorized based on their effect on UTRs from target genes. In a preferred embodiment, compounds can be categorized based on their ability to modulate the expression from a reporter gene operatively linked to a UTR. Categories of compounds can include, for example without limitation, compounds that modulate greater than or equal to 50% of the UTRs tested, compounds that modulate less than 50% modulation of the UTRs tested, compounds that modulate at least one UTR from a target gene at any concentration, compounds that modulate greater than or equal to 25% of the UTRs tested, compounds where the difference in modulation of at least one target UTR is greater than or equal to 25% of any other target UTR at any concentration tested, compounds where the difference in modulation of at least one target UTR is greater than or equal to 25% of any other UTR target for at least one concentration tested, and compounds with oddly-shaped dose-response curves for at least one target UTR tested. Compounds of the present invention can alternatively be classified based on the concentration where the compound is capable of modulating the expression from a reporter gene operatively linked to at least one target UTR.

In a preferred embodiment, most compounds lack UTR selectivity and similarly modulate the expression from a reporter gene operatively linked to at least one target UTR.

In a more preferred embodiment, most compounds lack UTR selectivity and similarly modulate the expression from a reporter gene operatively linked to at least one target UTR from at least four different target genes. In a most preferred embodiment, about 10 – about 50 compounds out of about 5,000 randomly chosen compounds will have pairwise  $IC_{50}$  ratios of 4-fold or more across at least four different target genes.

In a most preferred aspect, the activity or expression of a reporter gene is modulated without altering the activity of a control gene for general, indiscriminate translation activity. As used herein, indiscriminate translation activity refers to modulation in translation levels or activity that is random or unsystematic. One assay for modulation in general, indiscriminate translation activity uses a general translational inhibitor, for example puromycin, which is an inhibitor that causes release of nascent peptide and mRNA from actively translating ribosomes.

High-throughput screening can be done by exposing nucleic acid molecules of the present invention to a library of compounds and detecting gene expression with assays known in the art, including, for example without limitation, those described above. In one embodiment of the present invention, cancer cells, such as MCF-7 cells, expressing a nucleic acid molecule of the present invention are treated with a library of compounds. Percent inhibition of reporter gene activity can be obtained for all of the library compounds and can be analyzed using, for example without limitation, a scattergram generated by SpotFire® (SpotFire, Inc., Somerville, MA). The high-throughput screen can be followed by subsequent selectivity screens. In a preferred embodiment, a subsequent selectivity screen can include detection of reporter gene expression in cells expressing, for example, a reporter gene linked to a GEM or flanked by a 5' and 3' UTR of the same gene, either of which can contain a GEM of the present invention. In an alternative preferred embodiment, a subsequent selectivity screen can include detection of reporter gene expression in cells in the presence of a various concentrations of compounds.

Once a compound has been identified to modulate UTR-dependent expression of a target gene and preferably, the structure of the compound has been identified by the methods described in the present invention and well known in the art, the compounds are tested for biological activity in further assays and/or animal models. Further, a lead compound may be used to design congeners or analogs.

A wide variety of labels and conjugation techniques are known by those skilled in the art and can be used in various nucleic acid and amino acid assays. Methods for producing labeled hybridization or PCR probes for detecting sequences related to GEMs of the present invention include oligolabeling, nick translation, end-labeling, or PCR amplification using a labeled nucleotide. Suitable reporter molecules or labels which can be used for ease of detection include radionuclides, enzymes, and fluorescent, chemiluminescent, or chromogenic agents, as well as substrates, cofactors, inhibitors, magnetic particles, and the like.

*In vitro*

The present invention includes and provides for assays capable of screening for compounds capable of modulating gene expression. In a preferred aspect of the present invention, an assay is an *in vitro* assay. In a preferred aspect of the present invention, an *in vitro* assay that measures translation. In a preferred aspect of the present invention the *in vitro* assay includes a nucleic acid molecule of the present invention or a construct of the present invention.

In one embodiment, a reporter gene of the present invention can encode a fusion protein or a fusion protein comprising a domain that allows the expressed reporter gene to be bound to a solid support. For example, glutathione-S-transferase fusion proteins can be adsorbed onto glutathione sepharose beads (Sigma Chemical, St. Louis, MO) or glutathione derivatized microtiter plates, which are then combined with the compound or the compound and the non-adsorbed expressed reporter gene; the mixture is then incubated under conditions conducive to complex formation (e.g., at physiological conditions for salt and pH). Following incubation, the beads or microtiter plate wells are washed to remove any unbound components. Binding of the interactants can be determined either directly or indirectly, as described above. Alternatively, the complexes can be dissociated from the solid support before binding is determined.

Other techniques for immobilizing an expressed reporter gene or compound on a solid support also can be used in the screening assays of the invention. For example, either an expressed reporter gene or compound can be immobilized utilizing conjugation of biotin and streptavidin. Biotinylated expressed reporter genes or compounds can be prepared from biotin-NHS(N-hydroxysuccinimide) using techniques well known in the art (e.g.,

biotinylation kit, Pierce Chemicals, Rockford, IL) and immobilized in the wells of streptavidin-coated 96 well plates (Pierce Chemicals, Rockford, IL). Alternatively, antibodies which specifically bind to an expressed reporter gene or compound, but which do not interfere with a desired binding or catalytic site, can be derivatized to the wells of the plate. Unbound target or protein can be trapped in the wells by antibody conjugation.

Methods for detecting such complexes, in addition to those described above for the GST-immobilized complexes, include immunodetection of complexes using antibodies which specifically bind to an expressed reporter gene or compound, enzyme-linked assays which rely on detecting an activity of an expressed reporter gene, electrophoretic mobility shift assays (EMSA), and SDS gel electrophoresis under reducing or non-reducing conditions.

In one embodiment, translation of a reporter gene *in vitro* can be detected following the use of a reticulocyte lysate translation system, for example the TNT<sup>®</sup> Coupled Reticulocyte Lysate System (Promega Co., Madison, WI). In this aspect, for example, without limitation, RNA (100 ng) can be translated at 30° C in reaction mixtures containing 70% reticulocyte lysate, 20  $\mu$ M amino acids and RNase inhibitor (0.8 units/ $\mu$ l). After 45 minutes of incubation, 20  $\mu$ l of Lucite can be added and luminescence can be read on the View-Lux. Different concentrations of compounds can be added to the reaction in a final DMSO concentration of 2% and the EC<sub>50</sub> values calculated. Puromycin can be used as control for general indiscriminate translation inhibition. *In vitro* transcripts encoding a reporter gene linked to specific UTRs from target genes, including GAPDH, XIAP, TNF- $\alpha$ , and HIF-1 $\alpha$ , can also be used.

To study the influence of cell-type specific factors, capped RNA can be translated in translation extracts prepared from specialized cells or cancer cell lines, for example without limitation, HT1080 cells (a human fibrosarcoma cell line). Briefly, the cells can be washed with PBS and swollen in hypotonic buffer (10 mM Hepes, pH 7.4, 15 mM KCl, 1.5 mM Mg(OAc)<sub>2</sub>, 2 mM DTT and 0.5 mM Pefabloc (Pentapharm Ltd. Co., Switzerland) for 5 minutes on ice. The cells can be lysed using a Dounce homogenizer (100 strokes), and the extracts can be spun for 10 minutes at 10,000 x g. These clarified extracts can then be flash-frozen in liquid nitrogen and stored in aliquots at -70°C. The translation reaction can be capped RNA (50 ng) in a reaction mixture containing 60% clarified translation extract, 15  $\mu$ M total amino acids, 0.2 mg/ml Creatine phospho-kinase, which are all in 1X translation

buffer (15 mM Hepes, pH 7.4, 85 mM KOAc, 1.5 mM Mg(OAc)<sub>2</sub>, 0.5 mM ATP, 0.075 mM GTP, 18 mM creatine diphosphate and 1.5 mM DTT). After incubation of the translation reaction for 90 min at 37°C, activity of the protein encoded by the reporter gene can be detected. For activity of luciferase, encoded by the luciferase gene serving as the reporter gene, addition of 20 µl of LucLite® (Packard Instrument Co., Inc., Meriden, CT) can be used. Capped and uncapped RNAs can be synthesized *in vitro* using the T7 polymerase transcription kits (Ambion Inc., Austin, TX) and can be used in a similar *in vitro* system to study the influence of cell-type specific factors on translation.

#### *In vivo*

The present invention includes and provides for assays capable of screening for compounds capable of modulating gene expression. In a preferred aspect of the present invention, an assay is an *in vivo* assay. One preferred aspect of the present invention is an assay that measures translation. In a preferred embodiment of the present invention, an *in vivo* assay includes a nucleic acid molecule of the present invention or a construct of the present invention and can include the use of a cell or a cell or tissue within an organism. In a more preferred embodiment, an *in vivo* assay includes a nucleic acid molecule of the present invention present in a cell or a cell or tissue within an organism.

In another embodiment, *in vivo* translation of a reporter gene can be detected. In a preferred embodiment, a reporter gene is transfected into a cancer cell obtained from a cell line available at the (American Type Culture Collection (ATCC), Manassas, VA), for example HeLa, MCF-7, and COS-7, BT474. In a more preferred embodiment, a cancer cell has an altered genome relative to a similarly derived normal, primary cell, and the mammalian cancer cell proliferates under conditions where such a primary cell would not.

Screening for compounds that modulate reporter gene expression can be carried out in an intact cell. Any cell that comprises a reporter gene can be used in a cell-based assay system. A reporter gene can be naturally occurring in the cell or can be introduced using techniques such as those described above (see Cells and Organisms). In one embodiment, a cell line is chosen based on its expression levels of a naturally occurring protein, for example without limitation, VEGF, Her2, or survivin. Modulation of reporter gene expression by a compound can be determined *in vitro* as described above or *in vivo* as described below.



To detect expression of endogenous or heterologous proteins, a variety of protocols for detecting and measuring the expression of a reporter gene are known in the art. For example, Enzyme-Linked Immunosorbent Assays (ELISAs), western blots using either polyclonal or monoclonal antibodies specific for an expressed reporter gene, Fluorescence-Activated Cell Sorter (FACS), electrophoretic mobility shift assays (EMSA), or radioimmunoassay (RIA) can be performed to quantify the level of specific proteins in lysates or media derived from cells treated with the compounds. In a preferred embodiment, a phenotypic or physiological readout can be used to assess UTR-dependent activity of the target RNA in the presence and absence of the lead compound.

A wide variety of labels and conjugation techniques are known by those skilled in the art and can be used in various nucleic acid and amino acid assays. Methods for producing labeled hybridization or PCR probes for detecting sequences related to polynucleotides having a GEM of the present invention include oligolabeling, nick translation, end-labeling, or PCR amplification using a labeled nucleotide. Alternatively, sequences having a GEM of the present invention can be cloned into a vector for the production of a mRNA probe. Such vectors are known in the art, are commercially available, and can be used to synthesize RNA probes *in vitro* by addition of labeled nucleotides and an appropriate RNA polymerase such as T7, T3, or SP6. These procedures can be conducted using a variety of commercially available kits (Amersham Biosciences Inc., Piscataway, NJ; and Promega Co, Madison, WI). Suitable reporter molecules or labels which can be used for ease of detection include radionucleotides, enzymes, and fluorescent, chemiluminescent, or chromogenic agents, as well as substrates, cofactors, inhibitors, magnetic particles, and the like.

#### *Therapeutic Uses*

The present invention also provides for methods for treating, preventing or ameliorating one or more symptoms of a disease or disorder associated with the aberrant expression of a target gene, said method comprising administering to a subject in need thereof a therapeutically or prophylactically effective amount of a compound, or a pharmaceutically acceptable salt thereof, identified according to the methods described herein. In one embodiment, the target gene is aberrantly overexpressed. In another embodiment, the target gene is expressed at an aberrantly low level. In particular, the invention provides for a method of treating or preventing a disease or disorder or

ameliorating a symptom thereof, said method comprising administering to a subject in need thereof an effective amount of a compound, or a pharmaceutically acceptable salt thereof, identified according to the methods described herein, wherein said effective amount increases the expression of a target gene beneficial in the treatment or prevention of said disease or disorder. The invention also provides for a method of treating or preventing a disease or disorder or ameliorating a symptom thereof, said method comprising administering to a subject in need thereof an effective amount of a compound, or a pharmaceutically acceptable salt thereof, identified according to the methods described herein, wherein said effective amount decreases the expression of a target gene whose expression is associated with or has been linked to the onset, development, progression or severity of said disease or disorder. In a specific embodiment, the disease or disorder is a proliferative disorder, an inflammatory disorder, an infectious disease, a genetic disorder, an autoimmune disorder, a cardiovascular disease, or a central nervous system disorder. In an embodiment wherein the disease or disorder is an infectious disease, the infectious disease can be caused by a fungal infection, a bacterial infection, a viral infection, or an infection caused by another type of pathogen.

In addition, the present invention also provides pharmaceutical compositions that can be administered to a patient to achieve a therapeutic effect. Pharmaceutical compositions of the invention can comprise, for example, ribozymes or antisense oligonucleotides, antibodies that specifically bind to a GEM of the present invention, or mimetics, activators, inhibitors of GEM activity, or a nucleic acid molecule of the present invention. The compositions can be administered alone or in combination with at least one other agent, such as stabilizing compound, which can be administered in any sterile, biocompatible pharmaceutical carrier, including, but not limited to, saline, buffered saline, dextrose, and water. The compositions can be administered to a patient alone, or in combination with other agents, drugs or hormones.

In addition to the active ingredients, these pharmaceutical compositions can contain suitable pharmaceutically-acceptable carriers comprising excipients and auxiliaries which facilitate processing of the active compounds into preparations which can be used pharmaceutically. Pharmaceutical compositions of the invention can be administered by any number of routes including, but not limited to, oral, intravenous, intramuscular, intra-arterial, intramedullary, intrathecal, intraventricular, transdermal, subcutaneous, intraperitoneal,

intranasal, parenteral, topical, sublingual, or rectal means. Pharmaceutical compositions for oral administration can be formulated using pharmaceutically acceptable carriers well known in the art in dosages suitable for oral administration. Such carriers enable the pharmaceutical compositions to be formulated as tablets, pills, dragees, capsules, liquids, gels, syrups, slurries, suspensions, and the like, for ingestion by the patient.

Pharmaceutical preparations for oral use can be obtained through combination of active compounds with solid excipient, optionally grinding a resulting mixture, and processing the mixture of granules, after adding suitable auxiliaries, if desired, to obtain tablets or dragee cores. Suitable excipients are carbohydrate or protein fillers, such as sugars, including lactose, sucrose, mannitol, or sorbitol; starch from corn, wheat, rice, potato, or other plants; cellulose, such as methyl cellulose, hydroxypropylmethyl-cellulose, or sodium carboxymethylcellulose; gums including arabic and tragacanth; and proteins such as gelatin and collagen. If desired, disintegrating or solubilizing agents can be added, such as the cross-linked polyvinyl pyrrolidone, agar, alginic acid, or a salt thereof, such as sodium alginate.

Pharmaceutical preparations that can be used orally include push-fit capsules made of gelatin, as well as soft, sealed capsules made of gelatin and a coating, such as glycerol or sorbitol. Push-fit capsules can contain active ingredients mixed with fillers or binders, such as lactose or starches, lubricants, such as talc or magnesium stearate, and, optionally, stabilizers. In soft capsules, the active compounds can be dissolved or suspended in suitable liquids, such as fatty oils, liquid, or liquid polyethylene glycol with or without stabilizers.

Pharmaceutical formulations suitable for parenteral administration can be formulated in aqueous solutions, preferably in physiologically compatible buffers such as Hanks' solution, Ringer's solution, or physiologically buffered saline. Aqueous injection suspensions can contain substances that increase the viscosity of the suspension, such as sodium carboxymethyl cellulose, sorbitol, or dextran. Additionally, suspensions of the active compounds can be prepared as appropriate oily injection suspensions. Suitable lipophilic solvents or vehicles include fatty oils such as sesame oil, or synthetic fatty acid esters, such as ethyl oleate or triglycerides, or liposomes. Non-lipid polycationic amino polymers also can be used for delivery. Optionally, the suspension also can contain suitable stabilizers or agents that increase the solubility of the compounds to allow for the preparation of highly concentrated solutions. For topical or nasal administration, penetrants appropriate to the

particular barrier to be permeated are used in the formulation. Such penetrants are generally known in the art.

The pharmaceutical compositions of the present invention can be manufactured in a manner that is known in the art, e.g., by methods of conventional mixing, dissolving, granulating, dragee-making, levigating, emulsifying, encapsulating, entrapping, or lyophilizing processes. The pharmaceutical composition can be provided as a salt and can be formed with many acids, including but not limited to, hydrochloric, sulfuric, acetic, lactic, tartaric, malic, succinic, etc. Salts tend to be more soluble in aqueous or other protonic solvents than are the corresponding free base forms. In other cases, the preferred preparation can be a lyophilized powder which can contain any or all of the following: 1-50 mM histidine, 0.1%-2% sucrose, and 2-7% mannitol, at a pH range of 4.5 to 5.5, that is combined with buffer prior to use. Further details on techniques for formulation and administration can be found in the latest edition of *Remington's Pharmaceutical Sciences* (Maack Publishing Co., Easton, Pa.). After pharmaceutical compositions have been prepared, they can be placed in an appropriate container and labeled for treatment of an indicated condition. Such labeling would include amount, frequency, and method of administration.

#### *Determination of a Therapeutically Effective Dose*

A therapeutically effective dose refers to that amount of active ingredient that increases or decreases reporter gene activity relative to reporter gene activity that occurs in the absence of the therapeutically effective dose. For any compound, the therapeutically effective dose can be estimated initially either in cell culture assays or in animal models, usually mice, rabbits, dog, or pigs. The animal model also can be used to determine the appropriate concentration range and route of administration. Such information can then be used to determine useful doses and routes for administration in humans.

Therapeutic efficacy and toxicity, e.g., ED<sub>50</sub> (the dose therapeutically effective in 50% of the population) and LD<sub>50</sub> (the dose lethal to 50% of the population), can be determined by standard pharmaceutical procedures in cell cultures or experimental animals. The dose ratio of toxic to therapeutic effects is the therapeutic index, and it can be expressed as the ratio, LD<sub>50</sub>/ED<sub>50</sub>.

Pharmaceutical compositions that exhibit large therapeutic indices are preferred. The data obtained from cell culture assays and animal studies is used in formulating a range of

dosage for human use. The dosage contained in such compositions is preferably within a range of circulating concentrations that include the ED<sub>50</sub> with little or no toxicity. The dosage varies within this range depending upon the dosage form employed, sensitivity of the patient, and the route of administration.

The exact dosage will be determined by the practitioner, in light of factors related to the subject that requires treatment. Dosage and administration are adjusted to provide sufficient levels of the active ingredient or to maintain the desired effect. Factors that can be taken into account include the severity of the disease state, general health of the subject, age, weight, and gender of the subject, diet, time and frequency of administration, drug combination(s), reaction sensitivities, and tolerance/response to therapy. Long-acting pharmaceutical compositions can be administered every 3 to 4 days, every week, or once every two weeks depending on the half-life and clearance rate of the particular formulation.

Normal dosage amounts can vary from 0.1 to 100,000 micrograms, up to a total dose of about 1 g, depending upon the route of administration. Guidance as to particular dosages and methods of delivery is provided in the literature and generally available to practitioners in the art. Those skilled in the art will employ different formulations for nucleotides than for proteins or their inhibitors. Similarly, delivery of polynucleotides or polypeptides will be specific to particular cells, conditions, locations, etc.

If the reagent is a single-chain antibody, polynucleotides encoding the antibody can be constructed and introduced into a cell either *ex vivo* or *in vivo* using well-established techniques including, but not limited to, transferrin-polycation-mediated DNA transfer, transfection with naked or encapsulated nucleic acids, liposome-mediated cellular fusion, intracellular transportation of DNA-coated latex beads, protoplast fusion, viral infection, electroporation, "gene gun," and DEAE- or calcium phosphate-mediated transfection.

Effective *in vivo* dosages of an antibody are in the range of about 5 µg to about 50 µg/kg, about 50 µg to about 5 mg/kg, about 100 µg to about 500 µg/kg of patient body weight, and about 200 to about 250 µg/kg of patient body weight. For administration of polynucleotides encoding single-chain antibodies, effective *in vivo* dosages are in the range of about 100 ng to about 200 ng, 500 ng to about 50 mg, about 1 µg to about 2 mg, about 5 µg to about 500 µg, and about 20 µg to about 100 µg of DNA.

If the expression product is mRNA, the reagent is preferably an antisense

oligonucleotide or a ribozyme. Polynucleotides that express antisense oligonucleotides or ribozymes can be introduced into cells by a variety of methods, as described above.

Preferably, a reagent reduces expression of a reporter gene or the activity of a reporter gene by at least about 10, preferably about 50, more preferably about 75, 90, or 100% relative to the absence of the reagent. Alternatively, a reagent increases expression of a reporter gene or the activity of a reporter gene by at least about 10, preferably about 50, more preferably about 75, 90, or 100% relative to the absence of the reagent. The effectiveness of the reagent or mechanism chosen to modulate the level of expression of a reporter gene or the activity of a reporter gene can be assessed using methods well known in the art, such as hybridization of nucleotide probes to reporter gene-specific mRNA, quantitative RT-PCR, immunologic detection of an expressed reporter gene, or measurement of activity from an expressed reporter gene.

In any of the embodiments described above, any of the pharmaceutical compositions of the invention can be administered in combination with other appropriate therapeutic agents. Selection of the appropriate agents for use in combination therapy can be made by one of ordinary skill in the art, according to conventional pharmaceutical principles. The combination of therapeutic agents can act synergistically to effect the treatment or prevention of the various disorders described above. Using this approach, one may be able to achieve therapeutic efficacy with lower dosages of each agent, thus reducing the potential for adverse side effects.

Any of the therapeutic methods described above can be applied to any subject in need of such therapy, including, for example, mammals such as dogs, cats, cows, horses, rabbits, monkeys, and most preferably, humans.

#### *Administration of a Therapeutically Effective Dose*

A reagent which affects translation, either *in vitro* or *in vivo*, can be administered to a human cell to specifically reduce translational activity of a specific gene. In a preferred embodiment, the reagent preferably binds to a 5' UTR of a gene. In an alternate embodiment, the present invention the reagent preferably binds to a GEM of the present invention. In a preferred embodiment, the reagent is a compound. For treatment of human cells *ex vivo*, an antibody can be added to a preparation of stem cells which have been

removed from the body. The cells can then be replaced in the same or another human body, with or without clonal propagation, as is known in the art.

In one embodiment, the reagent is delivered using a liposome. Preferably, the liposome is stable in the animal into which it has been administered for at least about 30 minutes, more preferably for at least about 1 hour, and even more preferably for at least about 24 hours. A liposome comprises a lipid composition that is capable of targeting a reagent, particularly a polynucleotide, to a particular site in an animal, such as a human. Preferably, the lipid composition of the liposome is capable of targeting to a specific organ of an animal, such as the lung, liver, spleen, heart brain, lymph nodes, and skin.

A liposome useful in the present invention comprises a lipid composition that is capable of fusing with the plasma membrane of the targeted cell to deliver its contents to the cell. Preferably, the transfection efficiency of a liposome is about 0.5  $\mu\text{g}$  of DNA per 16 nmole of liposome delivered to about  $10^6$  cells, more preferably about 1.0  $\mu\text{g}$  of DNA per 16 nmole of liposome delivered to about  $10^6$  cells, and even more preferably about 2.0  $\mu\text{g}$  of DNA per 16 nmol of liposome delivered to about  $10^6$  cells. Preferably, a liposome is between about 100 and 500 nm, more preferably between about 150 and 450 nm, and even more preferably between about 200 and 400 nm in diameter.

Suitable liposomes for use in the present invention include those liposomes standardly used in, for example, gene delivery methods known to those of skill in the art. More preferred liposomes include liposomes having a polycationic lipid composition and/or liposomes having a cholesterol backbone conjugated to polyethylene glycol. Optionally, a liposome comprises a compound capable of targeting the liposome to a particular cell type, such as a cell-specific ligand exposed on the outer surface of the liposome.

Complexing a liposome with a reagent such as an antisense oligonucleotide or ribozyme can be achieved using methods that are standard in the art (see, for example, U.S. Pat. No. 5,705,151). Preferably, from about 0.1  $\mu\text{g}$  to about 10  $\mu\text{g}$  of polynucleotide is combined with about 8 nmol of liposomes, more preferably from about 0.5  $\mu\text{g}$  to about 5  $\mu\text{g}$  of polynucleotides are combined with about 8 nmol liposomes, and even more preferably about 1.0  $\mu\text{g}$  of polynucleotides is combined with about 8 nmol liposomes.

In another embodiment, antibodies can be delivered to specific tissues in vivo using receptor-mediated targeted delivery. Receptor-mediated DNA delivery techniques are taught

in, for example, Findeis et al. Trends in Biotechnol. 11, 202-05 (1993); Chiou et al., *Gene Therapeutics: Methods And Applications Of Direct Gene Transfer* (J. A. Wolff, ed.) (1994); Wu & Wu, J. Biol. Chem. 263, 621-24 (1988); Wu et al., J. Biol. Chem. 269, 542-46 (1994); Zenke et al., Proc. Natl. Acad. Sci. U.S.A. 87, 3655-59 (1990); Wu et al., J. Biol. Chem. 266, 338-42 (1991).

#### *Diagnostic Methods*

Agents of the present invention can also be used in diagnostic assays for detecting diseases and abnormalities or susceptibility to diseases and abnormalities related to the presence of mutations in the nucleic acid sequences that encode a GEM of the present invention. For example, differences can be determined between the cDNA or genomic sequence encoding a GEM in individuals afflicted with a disease and in normal individuals. If a mutation is observed in some or all of the afflicted individuals but not in normal individuals, then the mutation is likely to be the causative agent of the disease.

For example, the direct DNA sequencing method can reveal sequence differences between a reference gene and a gene having mutations. In addition, cloned DNA segments can be employed as probes to detect specific DNA segments. The sensitivity of this method is greatly enhanced when combined with PCR. For example, a sequencing primer can be used with a double-stranded PCR product or a single-stranded template molecule generated by a modified PCR. The sequence determination is performed by conventional procedures using radiolabeled nucleotides or by automatic sequencing procedures using fluorescent tags.

Moreover, for example, genetic testing based on DNA sequence differences can be carried out by detection of alteration in electrophoretic mobility of DNA fragments in gels with or without denaturing agents. Small sequence deletions and insertions can be visualized, for example, by high-resolution gel electrophoresis. DNA fragments of different sequences can be distinguished on denaturing formamide gradient gels in which the mobilities of different DNA fragments are retarded in the gel at different positions according to their specific melting or partial melting temperatures (see, e.g., Myers et al., Science 230, 1242, 1985). Sequence changes at specific locations can also be revealed by nuclease protection assays, such as RNase and S1 protection or the chemical cleavage method (e.g., Cotton et al., Proc. Natl. Acad. Sci. USA 85, 4397-4401, 1985). Thus, the detection of a specific DNA sequence can be performed by methods such as hybridization, RNase protection, chemical



cleavage, direct DNA sequencing or the use of restriction enzymes and Southern blotting of genomic DNA. In addition to direct methods such as gel-electrophoresis and DNA sequencing, mutations can also be detected by *in situ* analysis.

Altered levels of a GEM of the present invention can also be detected in various tissues. For example, one or more genes having a GEM can be detected by assays used to detect levels of particular nucleic acid sequence, such as Southern hybridization, northern hybridization, and PCR. Alternatively, assays can be used to detect levels of a reporter polypeptide regulated by a GEM or of a polypeptide encoded by a gene having a GEM. Such assays are well known to those of skill in the art and include radioimmunoassays, competitive binding assays, western blot analysis, and ELISA assays. A sample from a subject, such as blood or a tissue biopsy derived from a host, may be the material on which these assays are conducted.

Having now generally described the invention, the same will be more readily understood through reference to the following examples that are provided by way of illustration, and are not intended to be limiting of the present invention, unless specified.

Each periodical, patent, and other document or reference cited herein is herein incorporated by reference in its entirety.

#### Examples

*Example 1. Identification of compounds that specifically inhibit reporter gene expression post-transcriptionally.*

A monocistronic reporter construct (pLuc/veg5'+3'UTR) is under the transcriptional control of the CMV promoter and contains the VEGF 5' UTR driving the luciferase reporter upstream of the VEGF 3'UTR. Stable cell lines are generated by transfecting 293 cells with the pLuc/veg5'+3'UTR construct. A stable cell line is cultured under hygromycin B selection to create clonal cell lines consistent with protocols well known in the art. After two weeks of selection, clonal cell lines are screened for luciferase activity. The luciferase activity of several clonal cell lines (hereafter "Clones") are compared and normalized against total protein content. Clones are maintained under hygromycin B selection for more than three months with intermittent monitoring of luciferase activity. Clones are stable and maintain a high level of luciferase expression. Many Clones, for example, about twenty, may be compared to each other with respect to luciferase activity. In comparison to Clones B9,

D3, and H6, clone B9 exhibits the highest level of luciferase activity. In addition, semi-quantitative PCR analysis is performed, and the results indicate that multiple copies of the reporter are integrated per cell. Particular parameters for Clones are studied prior to selection for use in post-transcriptional, high-throughput screening (HTS). Relevant parameters for HTS include, but are not limited to, cell number, incubation time, DMSO concentration, and volume of substrate.

Chemical libraries in excess of 150,000 compounds are screened by HTS with a Clone containing the monocistronic reporter construct, pLuc/veg5' + 3'UTR. Screens are performed in duplicate with each molecule at a single concentration of about 7.5  $\mu$ M. Bright-Glow (Promega Co., Madison, WI) is used as a substrate to measure firefly luciferase activity. Active compounds are identified by reporting the average percent inhibition of the duplicate compounds followed by rejecting those compounds that did not provide satisfactory reproducibility. The average percent inhibition of compounds that provide satisfactory reproducibility is within a range of about 10%, about 25% or about 35% in the duplicate compounds. Data is analyzed as a normal distribution, which is apparent from graphical and statistical analysis of skewness and kurtosis. Hits are then reported at about a 99% confidence level, usually representing a selection of 3 standard deviations from the mean, or a hit lower limit of observed inhibition about equal to 50%. These selection criteria result in a hit rate of about 1%.

Certain compounds that are identified through the HTS-screening tier by screening with clone B9 modulate hypoxia-inducible endogenous VEGF expression. Endogenous VEGF protein levels are monitored by an ELISA assay (R&D Systems, Minneapolis, MN). HeLa cells are used to evaluate hypoxia-inducible expression. HeLa cells demonstrate about a three- to five-fold hypoxia-inducible window as compared to normoxic conditions (about 1000 - about 1500 pg/ml under hypoxia compared to about 200 – about 400 pg/ml under normoxia). Cells are cultured overnight to 48 hrs under hypoxic conditions (about 1% O<sub>2</sub>, about 5% CO<sub>2</sub>, and balanced with nitrogen) in the presence or absence of compounds. The conditioned media is assayed by ELISA. The concentration of VEGF is calculated from the standard ELISA curve of each assay. The assays are performed in duplicate at a compound concentration of about 7.5  $\mu$ M. A threshold of about 50% inhibition for a compound is selected as a criterion for further investigation. Further evaluation of about 100 to about 150 compounds is conducted from about 700 to about 800 initial HTS hits. The activity of the

identified compounds is confirmed by repeating the experiments described above. The identified compounds are then acquired as dry powders and analyzed further. The purity and molecular weight of the identified compounds are confirmed by LC-MS.

A dose-response analysis is performed using the ELISA assay and conditions described above. The conditions for the dose-response ELISA are analogous to those described above. A series of seven different serially-diluted concentrations are analyzed. In parallel, a dose-response cytotoxicity assay is performed under the same conditions as the ELISA to ensure that the inhibition of VEGF expression is not due to cytotoxicity. Dose-response curves are plotted using percentage inhibition versus log of concentration of the compound.

For each compound, the maximal inhibition is set as 100% and the minimal inhibition is set as 0% to generate  $EC_{50}$  and  $CC_{50}$  values. An identified compound from HTS shows a sigmoidal curve over a compound concentration range from about  $10^{-1}$  nM to about  $10^4$  nM when plotted as the log of concentration against the percent inhibition of VEGF expression on the y-axis. The same identified compound from HTS shows a convex curve over the same compound concentration range plotted against the percent of cytotoxicity. The ELISA  $EC_{50}$  (50% inhibition of VEGF expression) for this particular compound is about 7 nM, while its  $CC_{50}$  (50% cytotoxicity) is greater than about 2000 nM. Subsets of compounds that show similar efficacy/cytotoxicity windows are also identified.

The B9 cell line harbors the firefly luciferase reporter driven by the CMV promoter and flanked by the 5' and 3'UTRs of VEGF. Use of the B9 cell line with the HTS identifies compounds that specifically target the function of VEGF UTRs to modulate expression. Cell line B12 harbors the luciferase operably linked to control UTRs to replace the VEGF UTRs. Compounds that inhibit luciferase activity in both the B9 and B12 cell lines are general transcription and/or translation inhibitors or luciferase enzyme inhibitors. Several UTR specific compounds are identified in experiments with HTS identified compounds as described above. The dose-response curves of an identified compound show a sigmoidal curve in B9 cells and a concave curve in B12 cells when the percent luciferase inhibition of each is plotted over a compound concentration range from about  $10^{-1}$  nM to about  $10^4$  nM on the x-axis. The difference between the two cell lines (B9 and B12) shows that inhibition of VEGF production by this compound is through the VEGF UTRs, i.e., by a post-transcriptional control mechanism. A control experiment is performed with a

general translation inhibitor, puromycin. No difference in inhibition of luciferase expression is observed with puromycin treatment in these two cell lines.

*Example 2. Characteristics of UTR-specific VEGF inhibitors*

All identified compounds are re-synthesized and shown by LC/MS and combustion analysis to be greater than 95% pure. Subsequently, the re-synthesized compounds are tested in the dose-response VEGF ELISA and luciferase assays are used to initially assess UTR specificity. All identified compounds that retain UTR specificity are defined as *bona fide* UTR-specific inhibitors of VEGF expression.

High-throughput screening using B9 cells, followed by endogenous VEGF ELISAs identified compounds that specifically inhibit hypoxia inducible VEGF expression for the treatment of ocular neovascular diseases and cancer. Compounds that target multiple angiogenesis factors (including VEGF) for the treatment of cancers are also identifiable. Several targets are used for these purposes, including TNF- $\alpha$ , FGF-2, G-CSF, IGF-1, PDGF, and HIF-1 $\alpha$ .

ELISA assays analyze levels of expression of these factors using commercially available kits from R&D Systems (Minneapolis, MN). UTR-specific HTS identified compounds are tested for their ability to inhibit the expression of a subset of these proteins, including FGF-2 and IGF-1 in HeLa cells. Identified compounds that are very potent inhibitors of VEGF production as assayed in HeLa cells have EC<sub>50</sub> values ranging from low nM to high nM. Treatment with a general translation inhibitor (puromycin) results in similar inhibition for all these cytokines, with EC<sub>50</sub> values ranging from about 0.2 to about 2  $\mu$ M.

Lead compounds are further characterized and optimized. Analogs are synthesized and identified compounds exhibit excellent potency in the VEGF ELISA assay (EC<sub>50</sub> values ranging from 0.5 nM to 50 nM). In another embodiment, an analog exhibits low nM potency. In an additional embodiment, several analogs are synthesized and a subset of identified compounds are very active (EC<sub>50</sub> values ranging from 1 nM to 50 nM) in the VEGF ELISA assay. Activity of a very potent analog is improved about 500-fold compared to its parent (EC<sub>50</sub> of 1 nM vs. 500 nM). Further characterization and optimization for selectivity and pharmaceutical properties (ADMET) of the most active compounds will develop a drug candidate(s) for clinical trials.

*Example 3. HIF 1 $\alpha$  UTR modulates reporter gene expression*

**Transient Transfections:**

The HIF-1  $\alpha$  reporter constructs pGEMS HIF-1 $\alpha$ 5F3, pGEMS HIF-1 $\alpha$ 5F and pGEMS HIF-1 $\alpha$ F3 and pGEMS F are each transfected in equal amounts into 293 and MCF7 cells using FuGENE<sup>TM</sup> 6 (Fugent, LLC) transfection reagent (F. Hoffmann-La Roche Ltd, Basel, Switzerland) according to the manufacturer's instructions. The plasmid phRL-CMV is co-transfected with each reporter to normalize for transfection efficiencies. After 24 hours, transfected cells are washed with PBS, washed again with new media, and placed either under normoxia or hypoxia for another 24 hours. At that time, cells are harvested and assayed for Renilla and Firefly luciferase activities using the Dual-Luciferase reporter assay system (Promega, Inc., Madison, WI) according to the manufacturer's instructions.

**DNA Transfection and Generation of Stable Cell Line:**

To generate a stable cell line, 293 cells are transiently transfected with pGEMS HIF-1 $\alpha$ 5F3 as described above. Forty-eight hours later, cells are trypsinized, counted and seeded (10 ml) in 10 cm petri dishes at a concentration of 5000 cells/ml. The next day, 200  $\mu$ g/ml hygromycin B is added to the culture media in order to select for cells in which the transiently transfected plasmid has stably integrated into the genome. Following ten to fourteen days of hygromycin B selection, individual hygromycin-resistant clones are expanded by transferring the cells from the petri dish to a single well in a twenty-four well plate using trypsin-soaked filter discs according to manufacturer's instructions. Individual cell lines are then selected for further studies based on firefly luciferase expression levels.

**Luciferase Assay:**

Firefly and Renilla luciferase activities or Firefly luciferase activity only are measured using the Dual Luciferase or the Luciferase reporter assay systems (Promega Inc., Madison, WI), according to the manufacturer's instructions respectively.

**Quality control of stable clones using RT-PCR:**

Total RNA is isolated from each stably transfected clone obtained using Trizol<sup>®</sup> reagent (Invitrogen Co., Carlsbad, CA) according to the manufacturer's instructions. RT-PCR is then used in order to confirm the presence of the correct-sized HIF-1  $\alpha$  UTRs in the firefly reporter mRNAs isolated from the stable clones. Either a HIF-1  $\alpha$  5' UTR forward primer and a luciferase 5' reverse primer (5' CTGCAACTCCGATAAATAACGCGCCCAACA 3', SEQ ID NO:1) or a luciferase 3'

forward primer (5' CGGGTACCGAAAGGTCTTACC 3', SEQ ID NO: 2) and the HIF-1  $\alpha$  3' UTR reverse primer are used to amplify the 5' and 3' ends of the mRNA, respectively, from reverse-transcribed RNA using random hexamers.

**Quantitation of luciferase reporter RNA using Real Time RT-PCR:**

Luciferase reporter mRNA levels from all stable clones obtained are quantified using TaqMan<sup>®</sup> Real Time RT-PCR (Applied Biosystems, Foster City, CA) according to the manufacturer's instructions. The following firefly luciferase specific primers and probe are used: FLuc F (5' TTCTTCATAGTTGACCGCTTGA 3', SEQ ID NO: 3), FLuc R (5' GTCATCGTCGGGAAGACCT 3', SEQ ID NO: 4) and FLuc probe (5' 6FAM-CGATATTGTTACAACAACCAACATCTTCG-TAMRA 3', SEQ ID NO: 5 labeled with 6FAM at the 5' end and TAMRA at the 3' end). The luciferase reporter mRNA levels are normalized to actin mRNA levels using a commercially available actin-specific primers/probe set (Applied Biosystems, Foster City, CA).

**High Throughput Screening:**

High throughput screening ("HTS") for compounds that inhibit untranslated region-dependent expression of HIF-1  $\alpha$  is accomplished using stable cell line generated as described above. A 293 cell line contains stably integrated copies of the firefly luciferase gene flanked by both the 5' and 3' UTRs of HIF-1  $\alpha$ . The selected stable cell line is then used in a cell-based assay that has been optimized for cell number and percentage DMSO used for HTS.

Screening of compounds is accomplished within a week at a rate of 140 384-well plates per day. Each 384-well plate contains a standard puromycin titration curve that is used as a reference to calculate percent inhibition and the statistical significance of the data points generated in the assay. This curve is set-up in columns 3 and 4 of the 384-well plate and starts at a puromycin concentration of 20  $\mu$ M that is then serially diluted 2-fold down to 0.078  $\mu$ M and plated in quadruplicate. Columns 1 and 2 contain 16 standards each of a positive control consisting of cells in 0.5% DMSO and a negative control consisting of cells in 20  $\mu$ M puromycin. The difference between the two controls is used as the window to calculate the percentage of inhibition of luciferase expression in the presence of a compound. Columns 5 through 24 contain compounds from a library of small molecules.

HIF-1  $\alpha$  stable cells at a  $\sim 70$  % confluency are used for HTS. Briefly, the cells are dislodged from the flask with 4 ml of 0.25 % trypsin-EDTA (Gibco BRL, cat no. 25200-056) and diluted to 10 ml with non-selection media. This is repeated for all fourteen flasks and the cells are combined, passed through a filter, counted and diluted to a concentration of  $1.3 \times 10^5$  cells/ml. Cells in a volume of 38  $\mu$ l are added to each well containing 2  $\mu$ l of compound from a small molecule library to yield a final compound concentration of 7.5  $\mu$ M (3.75 mg/ml) in 0.5 % DMSO. The puromycin standard curve also contains 0.5 % DMSO. The compound-treated cells are incubated overnight (approximately 16 hours) under normoxic conditions and 37° C in 5 % CO<sub>2</sub>. To monitor firefly luciferase activity, SteadyLite HTS (PerkinElmer Life and Analytical Sciences, Inc., Boston, MA) is prepared according to manufacturer's instructions and 20  $\mu$ l are added to each well. Firefly luciferase activity in each well is detected with the ViewLux™ 1430 ultraHTS Microplate Imager (PerkinElmer Life and Analytical Sciences, Inc., Boston, MA). All data obtained is uploaded into Activity Base for calculations and statistical analyses of the percentages of inhibition of luciferase activity.

*Example 4. A preferred construct of the present invention*

A high-level expression vector, pcDNA™3.1/Hygro (Invitrogen Corp., Carlsbad, CA), is prepared as follows. In a pcDNA™3.1/Hygro vector, the untranslated regions (UTRs) and restriction sites associated with cloning, expressing, or cloning and expressing a gene of interest or a reporter gene are removed or replaced.

Certain UTRs and restriction sites are native to high-copy mammalian expression vectors. A vector without UTRs and restriction sites is prepared as follows. Deletion mutagenesis is undertaken to remove UTRs and restriction sites from commercially-available vector, pcDNA™3.1/Hygro (Invitrogen Corp., Carlsbad, CA). The vector is constructed to remove a region that starts at the putative transcription start site of a UTR found upstream of the cloning site and continues in the 3' direction to the *Hind III* restriction site at the multiple cloning site of pcDNA™3.1/Hygro (Invitrogen Corp., Carlsbad, CA). The nucleic acid sequence removed is SEQ ID NO: 23 (5'- AGAGAACCCA CTGCTTACTG GCTTATCGAA ATTAATACGAC TCACTATAGG GAGACCCAAGC TGGCTAGCGT TTAAGCTTA - 3'). As such, UTRs that are native to the vector and heterologous to the target gene are removed. In pcDNA™3.1/Hygro (Invitrogen Corp., Carlsbad, CA), the UTR removed is from the bovine growth hormone gene. Another nucleic acid sequence that is

removed is the UTR formed in the region starting at the *Xho I* site of pcDNA<sup>TM</sup>3.1/Hygro (Invitrogen Corp., Carlsbad, CA) continuing in the 3' direction and ending at the poly(A) tail, which in pcDNA<sup>TM</sup>3.1/Hygro (Invitrogen Corp., Carlsbad, CA) corresponding to the poly(A) tail from bovine growth hormone gene. By removing the nucleic acids from the *Xho I* site to the poly(A) tail, the 3' UTR native to the vector is removed, and the nucleic acid sequence that is removed is SEQ ID NO: 24. (5' - CTCGAGTCTA GAGGGCCCGT TTAAACCCGCT GATCAGCCTC GACTGTGGCC TTCTAGTTGCC AGCCATCTGTTG TTGTCCCCTC CCCCGTCCCTT CCTTGACCCT GGAAGGTGCC ACTCCCCTG TCCTTTCCT - 3').

A target UTR is cloned into the vector using a *Hind III* site and a *BamHI* site, which is downstream of the *Hind III* site. A target 5' UTR is inserted with a start codon upstream of the *BamHI* site. The reporter gene replaces the sequence between the *BamHI* site and a *Not I* site. Between the *Not I* site and a downstream *Xho I* site, the target 3' UTR is inserted with a stop codon downstream of the *Not I* site. The reporter gene is flanking and directly linked to the target 5' UTR and the target 3' UTR.



## WHAT IS CLAIMED:

1. A nucleic acid construct comprising a high-level mammalian expression vector, an intron, and a nucleic acid sequence encoding a reporter polypeptide, wherein said nucleic acid sequence encoding a reporter polypeptide is proximally linked to a target untranslated region (UTR).
2. The nucleic acid construct according to claim 1, wherein said intron is located within a 5' UTR.
3. The nucleic acid construct according to claim 1, wherein said intron is located within a 3' UTR.
4. The nucleic acid construct according to claim 1, wherein said intron is located within said nucleic acid sequence encoding a reporter polypeptide.
5. The nucleic acid construct according to claim 4, wherein said intron located within said nucleic acid sequence encoding a reporter polypeptide is spliced out during pre-mRNA processing.
6. The nucleic acid construct according to claim 1, wherein said nucleic acid sequence encoding a reporter polypeptide is directly linked to a target UTR.
7. A nucleic acid construct comprising a high-level mammalian expression vector and a nucleic acid sequence encoding a reporter polypeptide, wherein said nucleic acid sequence encoding a reporter polypeptide is directly linked to one or more target UTRs.
8. The nucleic acid molecule according to claim 7, wherein said one or more target UTRs has an element selected from the group consisting of an iron response element ("IRE"), internal ribosome entry site ("IRES"), upstream open reading frame ("uORF"), male specific lethal element ("MSL-2"), G-quartet element, and 5'-terminal oligopyrimidine tract ("TOP").
9. The nucleic acid molecule according to claim 7, wherein said one or more target UTRs are from the same target gene.

10. The nucleic acid construct according to claim 7, wherein said high-level mammalian expression vector integrates randomly into the genome.
11. The nucleic acid construct according to claim 7, wherein said high-level mammalian expression vector integrates site-selectively into the genome.
12. The nucleic acid construct according to claim 7, wherein said high-level mammalian expression vector is an episomal mammalian expression vector.
13. The nucleic acid construct according to claim 7, wherein said reporter gene contains an intron.
14. The nucleic acid construct according to claim 7, wherein said one or more target UTRs contains an intron.
15. A nucleic acid molecule comprising a nucleic acid sequence encoding a reporter polypeptide directly linked to one or more target UTRs.
16. The nucleic acid molecule according to claim 15, wherein said nucleic acid sequence encoding a reporter polypeptide contains an intron.
17. A heterologous population of nucleic acid molecules, wherein said heterologous population comprises a reporter nucleic acid sequence, wherein said nucleic acid sequence encoding a reporter polypeptide is directly linked to one or more target UTRs.
18. The heterologous population of nucleic acid molecules according to claim 17, wherein said heterologous population is isolated from a stable cell line.
19. The heterologous population of nucleic acid molecules according to claim 17, wherein said heterologous population is produced *in vitro*.
20. The heterologous population of nucleic acid molecules according to claim 17, wherein said heterologous population is used to produce polypeptides *in vitro*.
21. The heterologous population of nucleic acid molecules according to claim 17, wherein said heterologous population of nucleic acid molecules each have a 5' cap.

22. The heterologous population of nucleic acid molecules according to claim 17, wherein said heterologous population is selected to exclude molecules with a 5' cap.
23. The heterologous population of nucleic acid molecules according to claim 17, wherein said heterologous population is poly-adenylated.
24. The heterologous population of nucleic acid molecules according to claim 17, wherein said heterologous population is not poly-adenylated.
25. A method of making a nucleic acid construct to screen for a compound comprising:
  - a) cloning a gene and a vector in said nucleic acid construct;
  - b) engineering said nucleic acid construct to prevent an expressed gene product from having a UTR not found in a target gene; and
  - c) directly linking a target UTR to said gene.
26. The method according to claim 25, further comprising: d) expressing said gene linked to a target UTR in an absence of a UTR not found in a target gene.
27. The method according to claim 25, wherein said gene encodes a reporter polypeptide.
28. The method according to claim 25, wherein a target UTR is a 5' UTR from a target gene and a second target UTR is a 3' UTR.
29. The method according to claim 28, wherein said first target UTR is from the same target gene as said second target UTR.
30. The method according to claim 28, wherein said first target UTR is from a different target gene as said second target UTR.
31. A method of screening for a compound that modulates expression of a polypeptide comprising:
  - a) maintaining a cell, wherein said cell has a nucleic acid molecule and said nucleic acid molecule comprises a gene encoding a reporter polypeptide and said reporter gene is flanked by a target 5' UTR and a target 3' UTR;
  - b) forming a UTR-complex in said cell;

- c) contacting a compound with said UTR-complex; and
  - d) detecting an effect of said compound on said UTR-complex.
32. The method according to claim 31, wherein said UTR-complex contains a gene expression modulator (GEM).
33. The method according to claim 31, wherein said detecting is selected from the group consisting of an RNA-protein interaction assay, mass spectroscopy, RNA footprint analysis, and an RNA subcellular localization assay.
34. The method according to claim 31, wherein said detecting is based on comparing the level of reporter polypeptide expressed by said cell in a presence of said compound relative to in an absence of said compound.
35. A method of screening *in vivo* for a compound that modulates UTR-dependent expression comprising:
- a) providing a cell having a nucleic acid construct comprising a high-expression, constitutive promoter upstream from a target 5' UTR, said target 5' UTR upstream from a nucleic acid sequence encoding a reporter polypeptide, and said nucleic acid sequence encoding a reporter polypeptide upstream from a target 3' UTR;
  - b) contacting said cell with a compound;
  - c) producing a nucleic acid molecule that contains a nucleic acid sequence encoding a reporter polypeptide and does not contain UTR not found in a target gene; and
  - d) detecting said reporter polypeptide.
36. A method of screening *in vitro* for a compound that modulates UTR-affected expression comprising:
- a) providing an *in vitro* translation system;
  - b) contacting said *in vitro* translation system with a compound and a nucleic acid molecule comprising a target 5' UTR, said target 5' UTR upstream from a nucleic acid sequence encoding a reporter polypeptide and said nucleic acid sequence encoding a reporter polypeptide upstream from a target 3' UTR, wherein said nucleic acid molecule

is in an absence of a UTR not found in a target gene; and

c) detecting said reporter polypeptide *in vitro*.

37. A method of expressing a nucleic acid molecule in a cell comprising:

a) providing a heterologous nucleic acid molecule to a cell, wherein said nucleic acid molecule comprises a nucleic acid sequence encoding a reporter polypeptide flanked by target UTRs in an absence of a UTR not found in a target gene; and

b) detecting said reporter polypeptide *in vivo*.

38. The method according to claim 37, wherein said heterologous nucleic acid molecule is produced by *in vitro* transcription.

39. The method according to claim 37, wherein said heterologous nucleic acid molecule is a synthetically produced RNA molecule.

40. The method according to claim 37, wherein said heterologous nucleic acid molecule is a small interfering RNA (siRNA) molecule.

41. A method of screening for a compound that modulates protein expression through a main ORF-independent, UTR-affected mechanism comprising:

a) growing a stable cell line having a reporter gene proximally linked to a target UTR;  
b) comparing said stable cell line in the presence of a compound relative to in an absence of said compound; and  
c) selecting for said compound that modulates protein expression through a main ORF-independent, UTR-affected mechanism.

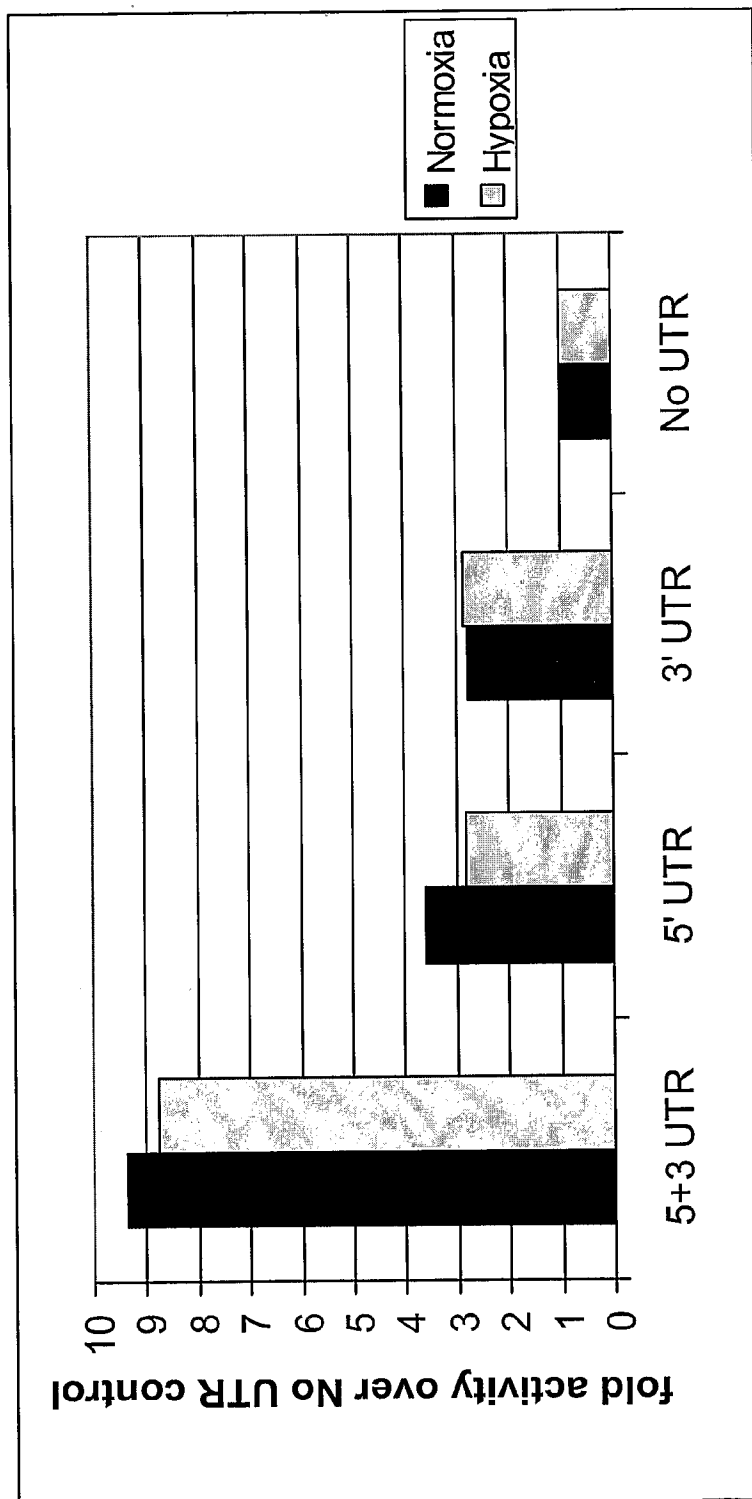
42. A method of screening for a compound that modulates protein expression through a main ORF-independent, UTR-affected mechanism comprising:

a) substituting in a cell a target gene with a reporter gene, wherein proximally linked target UTRs of said target gene remain intact and said cell is a differentiated cell;  
b) growing said cell line; and  
c) selecting for said compound that modulates protein expression of said reporter gene through a main ORF-independent, UTR-affected mechanism.

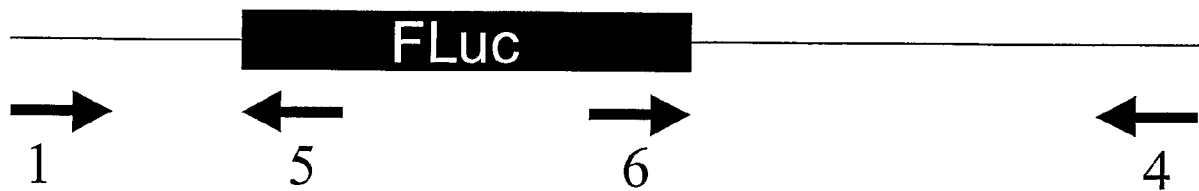
43. A method of screening for a compound that modulates protein expression through a UTR-affected mechanism comprising:
- a) growing a stable cell line having a reporter gene proximally linked to a target UTR, wherein said stable cell line mimics post-transcriptional regulation of a target gene found *in vivo*;
  - b) growing said stable cell line; and
  - c) selecting for said compound that modulates protein expression of said reporter gene through a UTR-affected mechanism.
44. The method according to claim 43, wherein the nucleic acid sequence of said target UTR is specific to said target gene in mammals.
45. The method according to claim 43, wherein the nucleic acid sequence of said target UTR is specific to said target gene in plants.
46. The method according to claim 43, wherein said target gene is an isoform.
47. The method according to claim 43, wherein said target gene contains a UTR also found in one or more different genes.
48. The method according to claim 47, wherein said target gene is indicative of a disease state.
49. The method according to claim 43, wherein said stable cell line is a cancer cell.
50. A method of screening for a compound that modulates protein expression through a UTR-affected mechanism comprising:
- a) growing a stable cell line having a reporter gene proximally linked to more than one target UTR;
  - b) comparing said stable cell line in the presence of a compound relative to in an absence of said compound, wherein said compound does not modulate UTR-dependent expression if only one target UTR is proximally linked to a reporter gene; and
  - c) selecting for said compound that modulates protein expression of said reporter gene through a UTR-affected mechanism.

51. The method according to claim 50, further comprising: d) comparing said modulation of UTR-dependent protein expression with a UTR not found in a target gene proximally linked to a reporter gene relative to modulation of UTR-dependent protein expression with a reporter gene flanked by a proximally linked target 5' UTR and a proximally linked target 3' UTR.
52. The method according to claim 50, further comprising: d) comparing modulation of UTR-dependent protein expression with a reporter gene having an intron relative to modulation of UTR-dependent protein expression to said reporter gene without said intron.
53. The method according to claim 50, wherein said compound affects a UTR-complex and said UTR-complex contains a protein selected from the group consisting of a small nuclear RNPs (snRNP), hnRNP proteins, mRNA proteins, splicing factors, ribosomal proteins, and translation-specific proteins that are non-ribosomal.
54. The method according to claim 53, wherein said UTR-complex does not include a protein selected from the group consisting of a non-regulatory ribosomal protein and a chromatin-associated protein.

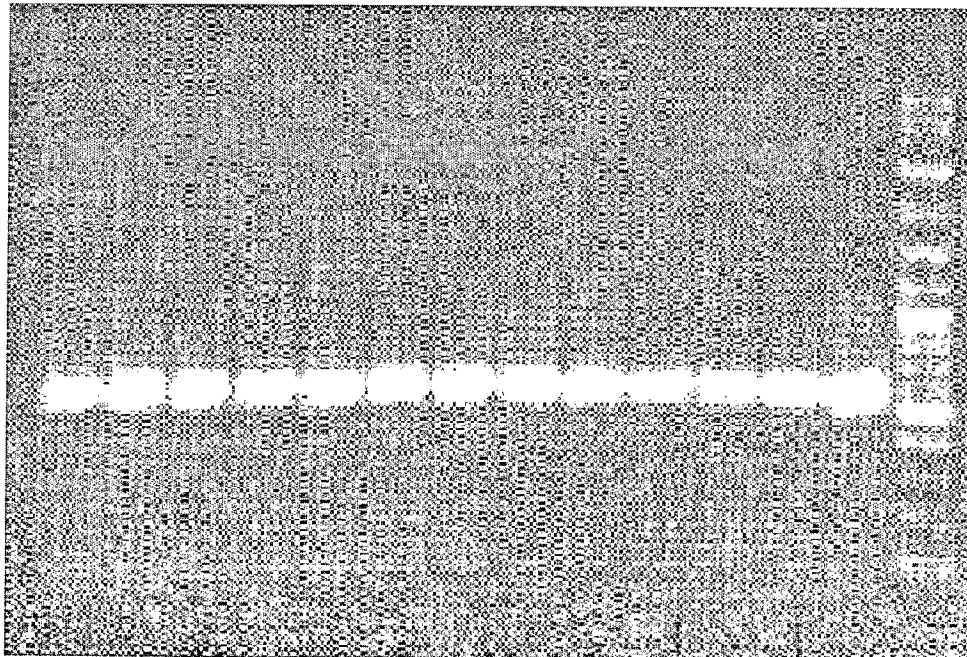
**FIG.1**





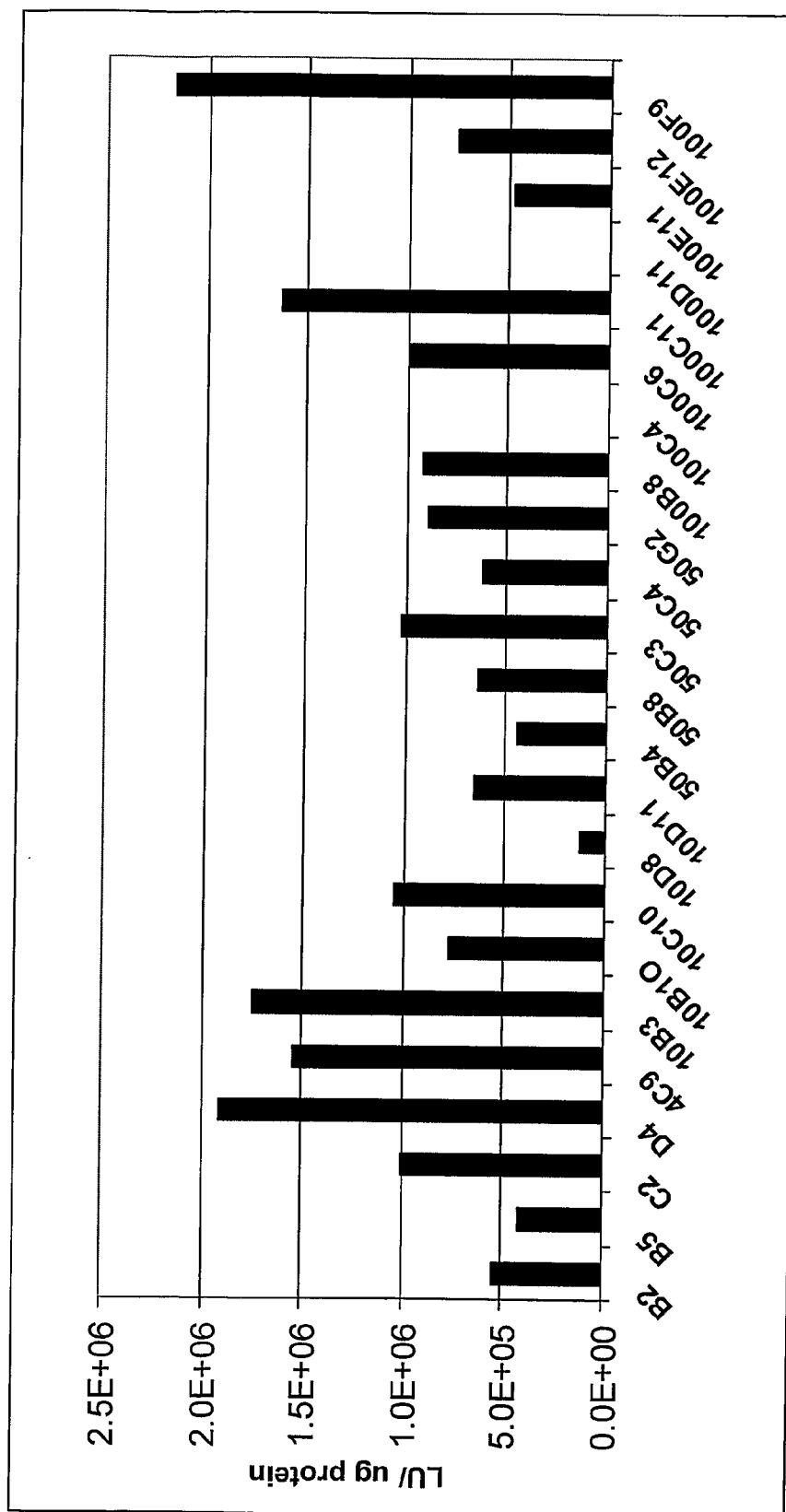
**FIG.2A****FIG.2B**

1 2 3 4 5 6 7 8 9 10 11 12 13 14

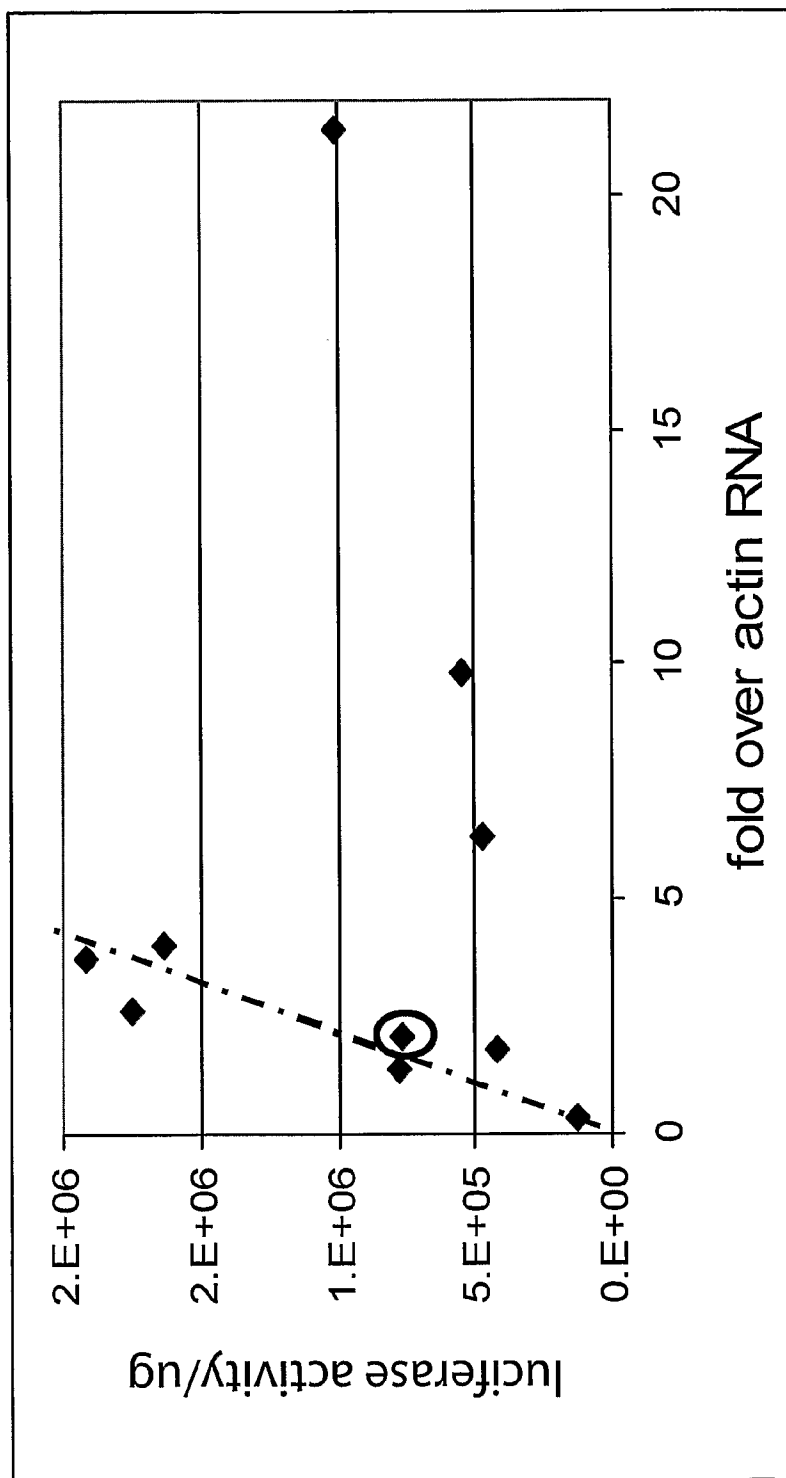


~0.5 kb

FIG. 3



**FIG. 4**



## SEQUENCE LISTING

<110> PTC Therapeutics, Inc.

<120> Methods and Agents for Screening for Compounds Capable of Modulating Gene Expression

<130> 19025.023

<140> To Be Determined

<141> 2004-08-16

<150> 10/895,393

<151> 2004-07-21

<160> 118

<170> PatentIn version 3.2

<210> 1

<211> 14

<212> DNA

<213> Artificial

<220>

<223> Description of Artificial Sequence: Motif

<220>

<221> misc\_feature

<222> 3, 7, 8, 11

<223> n = a, t, c, or g

<220>

<221> misc\_feature

<222> (7)..(8)

<223> This represents one form of the sequence as described, other forms described may have up to five nucleotides in this variable region

<400> 1  
ggntggngg ntgg 14

<210> 2

<211> 14

<212> DNA

<213> Artificial

<220>

<223> Description of Artificial Sequence: Motif

<220>

<221> misc\_feature

<222> 3, 4, 7, 8, 11, 12

<223> n = a, t, g or c

<220>

<221> misc\_feature

<222> (2)..(12)

<223> This represents one form of the sequence as described, other forms described have longer variable regions, typical is 2 - 10 nucleotides

<400> 2

ggnnnggnngg nngg

14

<210> 3

<211> 14

<212> DNA

<213> Artificial

<220>

<223> Description of Artificial Sequence: Motif

<220>

<221> misc\_feature

<222> 3, 4, 7, 8, 11, 12

<223> n = a, t, g, or c

<220>

<221> misc\_feature

<222> (2)..(12)

<223> This represents one form of the sequence as described, other forms described have longer variable regions, typical is 2 - 10 nucleotides

<400> 3

ggnnnggnngg nngg

14

<210> 4

<211> 19

<212> RNA

<213> Artificial

<220>

<223> Description of Artificial Sequence: Motif

<400> 4

ccccrcccuc uuccccaag

19

<210> 5

<211> 152

<212> DNA

<213> Homo sapiens

<400> 5

gcagaggacc agctaagagg gagagaagca actacagacc cccctgaaa acaaccctca

60

gacgccacat cccctgacaa gctgccaggc aggttctctt cctctcacat actgaccac 120  
 ggctccaccc tctctccct ggaaaggaca cc 152

<210> 6  
 <211> 792  
 <212> DNA  
 <213> Homo sapiens

<400> 6  
 tgaggaggac gaacatccaa ccttccaaa cgcctccct gcccaatcc ctttattacc 60  
 ccctccttca gacaccctca acctcttctg gctcaaaaag agaattgggg gcttagggtc 120  
 ggaaccaag cttagaactt taagcaaca gaccaccact tcgaaacctg ggattcagga 180  
 atgtgtggcc tgcacagtga attgctggca accactaaga attcaaatg gggcctccag 240  
 aactcactgg ggctacagc ttgatccct gacatctgga atctggagac caggagacct 300  
 ttggttcttg ccagaatgct gcaggacttg agaagacctc acctagaaat tgacacaagt 360  
 ggaccttagg ccttcctctc tccagatgtt tccagacttc cttgagacac ggagcccagc 420  
 cctcccatg gagccagctc cctctattta tgtttgcaact tgtgattatt tattatttat 480  
 ttattattta tttatttaca gatgaatgta tttatttggg agaccgggg atcctggggg 540  
 acccaatgta ggagctgct tggctcagac atgttttccg tgaaaacgga gctgaacaat 600  
 aggctgttcc catgtagccc cctggcctct gtgccttctt ttgattatgt tttttaaaat 660  
 atttatctga ttaagttgct taaacaatgc tgatttggtg accaactgct actcattgct 720  
 gagcctctgc tcccagggg agttgtgtct gtaatcgccc tactattcag tggcgagaaa 780  
 taaagtttgc tt 792

<210> 7  
 <211> 21  
 <212> RNA  
 <213> Artificial

<220>  
 <223> Description of Artificial Sequence: Motif

<400> 7  
 auuuuuuuu uuauuuuuu a 21

<210> 8  
 <211> 40  
 <212> DNA  
 <213> Homo sapiens

<400> 8  
kctggaggat gtggctgcag agcctgctgc tcttgggcac 40

<210> 9  
<211> 289  
<212> DNA  
<213> Homo sapiens

<400> 9  
gccggggagc tgctctctca tgaaacaaga gctagaaact caggatgggc atcttggagg 60  
gaccaagggg tggggccacag ccatgggtggg agtggcctgg acctgccctg ggccacactg 120  
acctgatac aggcattggca gaagaatggg aatattttat actgacagaa atcagtaata 180  
tttatatatt tatattttta aaatattttat ttattttatt atttaagttc atattccata 240  
tttattcaag atgtttttacc gtaataatta ttattaaaaa tatgcttct 289

<210> 10  
<211> 21  
<212> RNA  
<213> Artificial

<220>  
<223> Description of Artificial Sequence: Motif

<400> 10  
auuuauuuau uuauuuauuu a 21

<210> 11  
<211> 47  
<212> DNA  
<213> Homo sapiens

<400> 11  
atcactctct ttaatcacta ctcacattaa cctcaactcc tgccaca 47

<210> 12  
<211> 307  
<212> DNA  
<213> Homo sapiens

<400> 12  
taattaagtg cttcccactt aaaacatata aggccttota tttattttatt taaatatatta 60  
aattttatat ttattgttga atgtatgggt gctacctatt gtaactatta ttottaatct 120  
taaaactata aatatggatc ttttatgatt ctttttgtaa gccctagggg ctctaaaatg 180  
gtttacotta tttatcccaa aaatattttat tattatgttg aatgttaa atagtatcta 240  
tgtagattgg ttagtaaaac tatttaataa atttgataaa tataaaaaaa aaaaacaaaa 300

aaaaaaaa 307

<210> 13  
 <211> 15  
 <212> RNA  
 <213> Artificial

<220>  
 <223> Description of Artificial Sequence: Motif

<220>  
 <221> misc\_feature  
 <222> (1)..(15)  
 <223> n = a, t, g or c

<400> 13  
 nauuuauuuu uuuan 15

<210> 14  
 <211> 62  
 <212> DNA  
 <213> Homo sapiens

<400> 14  
 ttctgcccctc gagcccaccg ggaacgaaag agaagctcta tctgcctcc aggagcccag 60  
 ct 62

<210> 15  
 <211> 427  
 <212> DNA  
 <213> Homo sapiens

<400> 15  
 tagcatgggc acctcagatt gttgttggtta atgggcattc cttcttctgg tcagaaacct 60  
 gtccactggg cacagaactt atgttggttct ctatggagaa ctaaaagtat gagcgtagg 120  
 acactatttt aattattttt aatttattaa tatttaaata tgtgaagctg agttaattta 180  
 tgtaagtcac atttatattt ttaagaagta ccacttgaaa cattttatgt attagttttg 240  
 aaataataat ggaaagtggc tatgcagttt gaatatcctt tgtttcagag ccagatcatt 300  
 tcttggaag tgtaggctta cctcaaataa atggctaact tatacatatt tttaaagaaa 360  
 tatttatatt gtatttatat aatgtataaa tgggttttat accaataaat ggcattttta 420  
 aaaattc 427

<210> 16



<211> 15  
 <212> RNA  
 <213> Artificial

<220>  
 <223> Description of Artificial Sequence: Motif

<220>  
 <221> misc\_feature  
 <222> (1)..(15)  
 <223> n = a, t, g or c

<400> 16  
 nauuuauuuu uuuan 15

<210> 17  
 <211> 701  
 <212> DNA  
 <213> Homo sapiens

<400> 17  
 aagagctcca gagagaagtc gaggaagaga gagacggggt cagagagagc gcgcggggcgt 60  
 gcgagcagcg aaagcgacag gggcaaagtg agtgacctgc ttttgggggt gaccgccgga 120  
 gcgcggcgctg agccctcccc cttgggatcc cgcagctgac cagtcgcgct gacggacaga 180  
 cagacagaca ccgccccag cccagttac cacctcctcc ccggccggcg gcggacagtg 240  
 gacgcggcgg cgagccgcgg gcaggggccg gagcccgccc ccggaggcgg ggtggagggg 300  
 gtcggagctc gcggcgctgc actgaaactt ttcgtccaac ttctgggctg ttctcgcttc 360  
 ggaggagccg tggtcgcgcg gggggaagcc gagccgagcg gagccgcgag aagtgctagc 420  
 tcgggccggg aggagccgca gccggaggag ggggaggagg aagaagagaa ggaagaggag 480  
 agggggccgc agtggcgact cggcgctcgg aagccgggct catggacggg tgaggcgggc 540  
 gtgtgcgcag acagtgtctc agcgcgcgcg ctccccagcc ctggcccggc ctggggccgg 600  
 gaggaagagt agctcgccga ggcgccgagg agagcggggc gcccacagc ccgagccgga 660  
 gagggacgcg agccgcgcgc cccggtcggg cctccgaaac c 701

<210> 18  
 <211> 1892  
 <212> DNA  
 <213> Homo sapiens

<400> 18  
 tgagccgggc aggaggaagg agcctccctc agggtttcgg gaaccagatc tctctccagg 60  
 aaagactgat acagaacgat cgatacagaa accacgctgc cgccaccaca ccatcaccat 120

cgacagaaca gtccttaatc cagaaacctg aaatgaagga agaggagact ctgcgcagag 180  
 cactttgggt ccggagggcg agactccggc ggaagcattc ccgggcgggt gaccagcac 240  
 ggtccctctt ggaattggat tcgccatittt atttttcttg ctgctaaatc accgagcccg 300  
 gaagattaga gagttttatt tctgggattc ctgtagacac acccaccac atacatacat 360  
 ttatatatat atatattata tatatataaa aataaatatc tctattttat atatataaaa 420  
 tatatatatt ctttttttaa attaacagtg ctaatgttat tgggtgtcttc actggatgta 480  
 tttgactgct gtggacttga gttgggaggg gaatgttccc actcagatcc tgacagggaa 540  
 gaggaggaga tgagagactc tggcatgatc ttttttttgt cccacttggg ggggccaggg 600  
 tcctctcccc tgcccaagaa tgtgcaaggc cagggcattg gggcaaatat gaccagttt 660  
 tgggaacacc gacaaacca gccctggcgc tgagcctctc taccacaggc cagacggaca 720  
 gaaagacaaa tcacagggtc cgggatgagg acaccggctc tgaccaggag tttggggagc 780  
 ttcaggacat tgctgtgctt tggggattcc ctccacatgc tgcacgcgca tctcgcccc 840  
 aggggcactg cctggaagat tcaggagcct gggcggcctt cgcttactct cacctgcttc 900  
 tgagttgccc aggaggccac tggcagatgt cccggcgaag agaagagaca cattgttggg 960  
 agaagcagcc catgacagcg ccccttcctg ggaactcgccc tcctctctct cctgctccc 1020  
 ttcctggggg gcagcctaaa aggacctatg tctcacacc attgaaacca ctagtctctg 1080  
 cccccagga aacctggtg tgtgtgtgtg agtggttgac ctctctccat cccctggctc 1140  
 ttcccttccc ttcccgaggc acagagagac agggcaggat ccacgtgccc attgtggagg 1200  
 cagagaaaag agaaagtgtt ttatatacgg tacttattta atatcccttt ttaattagaa 1260  
 attagaacag ttaatttaat taaagagtag ggtttttttt cagtattctt ggtaattatt 1320  
 taatttcaac tatttatgag atgtatcttt tgctctctct tgctctctta tttgtacgg 1380  
 tttttgtata taaaattcat gttccaatc tctctctccc tgatcgggtga cagtcaactag 1440  
 cttatcttga acagatattt aattttgcta acactcagct ctgccctccc cgatcccctg 1500  
 gctccccagc acacattcct ttgaaagagg gtttcaatat acatctacat actatatata 1560  
 tattgggcaa cttgtatttg tgtgtatata tatatatata tgtttatgta tatatgtgat 1620  
 cctgaaaaaa taaacatcgc tattctgttt tttatatgtt caaacaaaac aagaaaaaat 1680  
 agagaattct acatactaaa tctctctcct tttttaattt taatatltgt tatcatttat 1740  
 ttattgggtgc tactgtttat ccgtaataat tgtggggaaa agatattaac atcacgtctt 1800

tgtctctagtg gcagtttttc gagatattcc gtagtacata tttattttta aacaacgaca 1860  
 aagaaataca gatatatctt aaaaaaaaaa aa 1892

<210> 19  
 <211> 249  
 <212> RNA  
 <213> Homo sapiens

<400> 19  
 ccgggcucan ggacggguga ggcgggcgug ugcgcagaca gugcuccagc gcgcgcgcuc 60  
 ccagagccug gcccgccuc gggcgggag gaagaguagc ugcgcgaggc gccgaggaga 120  
 gcgggccgcc ccacagcccg agccggagag ggacgcgagc gcgcgcgccc ggucgggccu 180  
 ccgaaaccan gaacuuucug cugucuuggg ugcuuaggag ccuugccuug cugcucuacc 240  
 uccaccaug 249

<210> 20  
 <211> 15  
 <212> RNA  
 <213> Artificial

<220>  
 <223> Description of Artificial Sequence: Motif

<220>  
 <221> misc\_feature  
 <222> (1)..(15)  
 <223> n = a, t, g or c

<400> 20  
 nauuuauuua uuuan 15

<210> 21  
 <211> 49  
 <212> DNA  
 <213> Homo sapiens

<400> 21  
 ccgccagatt tgaatcgagg gaccggttgg cagaggtggc ggcggcggc 49

<210> 22  
 <211> 1141  
 <212> DNA  
 <213> Homo sapiens

<400> 22  
 ggccctctggc cggagctgcc tgggtcccaga gtggctgcac cacttccagg gtttattccc 60

```

tggtgccacc agccttcctg tgggccctt agcaatgtct taggaaagga gatcaacatt 120
ttcaaattag atgtttcaac tgtgtcctg ttttgtcttg aaagtggcac cagagggtgct 180
tctgcctgtg cagcgggtgc tgcgtgtaac agtggctgct tctctctctc tctctctttt 240
ttgggggctc atttttgctg ttttgattcc cgggcttacc aggtgagaag tgaggaggga 300
agaaggcagt gtcccttttg ctagagctga cagctttgtt cgcgtgggca gagccttcca 360
cagtgaatgt gtctggacct catgttggtg aggctgtcac agtcctgagt gtggacttgg 420
cagggtgcctg ttgaatctga gctgcagggt ccttatctgt cacacctgtg cctcctcaga 480
ggacagtttt tttgttggtg tgtttttttg tttttttttt ttggtagatg catgacttgt 540
gtgtgatgag agaatggaga cagagtcctt ggctcctcta ctgtttaaca acatggcttt 600
cttattttgt ttgaattggt aattcacaga atagcacaaa ctacaattaa aactaagcac 660
aaagccattc taagtcattg gggaaacggg gtgaacttca ggtggatgag gagacagaa 720
agagtgatag gaagcgtctg gcagatactc cttttgccac tgctgtgtga ttagacaggc 780
ccagtgagcc gcggggcaca tgctggccgc tcctccctca gaaaaaggca gtggcctaaa 840
tcctttttta atgacttggc tcgatgctgt gggggactgg ctgggctgct gcaggccgtg 900
tgtctgtcag cccaaccttc acatctgtca cgttctccac acgggggaga gacgcagtcc 960
gccaggtcc ccgctttctt tggaggcagc agctcccgca gggctgaagt ctggcgtaag 1020
atgatggatt tgattcgccc tcctccctgt catagagctg cagggtggat tgttacagct 1080
tcgctggaaa cctctggagg tcatctcggc tgttcctgag aaataaaaag cctgtcattt 1140
c 1141

```

&lt;210&gt; 23

&lt;211&gt; 247

&lt;212&gt; DNA

&lt;213&gt; Homo sapiens

&lt;400&gt; 23

```

ccccggcgca gcgcggccgc agcagcctcc gcccccgca cgggtgtgagc gcccgacgcg 60
gccgaggcgg ccggagtccc gagctagccc cggcgccgc cgcgcgccag accggacgac 120
aggccacctc gtgcggctcc gcccgagtcc ccgcctcgcc gccaacgcca caaccaccgc 180
gcacggcccc ctgactccgt ccagtattga tcgggagagc cggagcgagc tcttcgggga 240
gcagcag 247

```

&lt;210&gt; 24

&lt;211&gt; 1716

&lt;212&gt; DNA

&lt;213&gt; Homo sapiens

&lt;400&gt; 24

```

tgaccacgga ggatagtatg agccctaaaa atccagactc tttcgatacc caggaccaag      60
ccacagcagg tectccatcc caacagccat gcccgcata gctcttagac ccacagactg      120
gttttgcaac gtttacaccg actagccagg aagtacttcc acctcgggca cattttggga      180
agttgcattc ctttgtcttc aaactgtgaa gcatttacag aaacgcattc agcaagaata      240
ttgtcccttt gagcagaaat ttatctttca aagaggtata ttgaaaaaa aaaaaaaaag      300
tatatgtgag gatttttatt gattggggat cttggagttt ttcattgtcg ctattgattt      360
ttacttcaat gggctcttcc aacaaggaag aagcttgctg gtagcacttg ctaccctgag      420
ttcatccagg cccaactgtg agcaaggagc acaagccaca agtcttccag aggatgcttg      480
attccagtg gttctgcttc aggtctccac tgcaaaacac taaagatcca agaaggcctt      540
catggcccca gcaggccgga tcgggtactgt atcaagtcac ggcaggtaga gtaggataag      600
ccactctgtc ccttcctggg caaagaagaa acggagggga tgaattcttc cttagactta      660
cttttgtaaa aatgtcccca cgggtacttac tccccactga tggaccagtg gtttccagtc      720
atgagcgtta gactgacttg tttgtcttcc attccattgt tttgaaactc agtatgccgc      780
ccctgtcttg ctgtcatgaa atcagcaaga gaggatgaca catcaaataa taactcggat      840
tccagccccc attggattca tcagcatttg gaccaatagc ccacagctga gaatgtggaa      900
tacctaagga taacaccgct tttgtttctg caaaaacgta tctcctaatt tgaggctcag      960
atgaaatgca tcaggtcctt tggggcatag atcagaagac tacaaaaatg aagctgctct     1020
gaaatctcct ttagccatca ccccaacccc ccaaaattag tttgtgttac ttatggaaga     1080
tagttttctc cttttacttc acttcaaaag ctttttactc aaagagtata tgttccctcc     1140
aggtcagctg cccccaaacc ccctccttac gctttgtcac aaaaaaagtg totctgcctt     1200
gagtcattct ttcaagcact tacagctctg gccacaacag ggcattttac aggtgcgaat     1260
gacagtagca ttatgagtag tgtgaattca ggtagtaaat atgaaactag ggtttgaaat     1320
tgataatgct ttcacaacat ttgcagatgt tttagaagga aaaaagttcc ttcttaaaat     1380
aattttctct caattggaag attggaagat tcagctagtt aggagcccat tttttcctaa     1440
tctgtgtgtg ccctgtaacc tgactgggta acagcagtc tttgtaaaca gtgttttaaa     1500
ctctcctagt caatatccac cccatccaat ttatcaagga agaaatggtt cagaaaaatat     1560

```

tttcagccta cagttatgtt cagtcacaca cacatacaaa atgttccttt tgccttttaaa 1620  
 gtaattttttg actcccagat cagtcagagc ccctacagca ttgttaagaa agtatttgat 1680  
 ttttgtctca atgaaaataa aactatatctc atttcc 1716

<210> 25  
 <211> 160  
 <212> DNA  
 <213> Homo sapiens

<400> 25  
 tataaaaagct gggccggcgc gggccgggccc attcgcgacc cggaggtgcg cgggcgcggg 60  
 cgagcagggt ctccgggtgg gcggcgcgac gccccgcgca ggctggaggc cgccgaggct 120  
 cgccatgcog ggagaactct aactccccc tggagtcggc 160

<210> 26  
 <211> 1306  
 <212> DNA  
 <213> Homo sapiens

<400> 26  
 tgaggcgcgc ggctgtggga cggccctggg ccagcctccg gcggggaccc agggagtggg 60  
 ttggggctgc cggatctcga ggcttgccca gaccgtgcga gccaggacta ggagattccg 120  
 gtgcctcctg aaagcctggc ctgctccgcg tgtccctcc cttcctctgc gccggacttg 180  
 gtgcgtctaa gatgaggggg ccaggcgggt gcttctccct gcgaggaggg gagaattctt 240  
 ggggctgagc tgggagccc gcaactctag tatttaggat aacttggtgc ttggaaatgc 300  
 aaactcaccg ctccaatgcc tactgagtag ggggagcaaa tcgtgccttg tcattttatt 360  
 tggaggtttc ctgcctcctt cccgaggcta cagcagacc ccatgagaga aggaggggag 420  
 caggcccgctg gaggaggggg gctcaggagg ctgagatccc gacaagcccg ccagccccag 480  
 ccgctcctcc acgcctgtcc ttagaaaggg gtggaacat agggacttgg ggcttggaac 540  
 ctaaggttgt tccctagttc tacatgaagg tggaggcttc tagttccacg cctctccac 600  
 ctccctccgc acacacccca ccagcctgc tataggctgg ctttcccttg gggctggaac 660  
 tcaactgcgat ggggtcacca ggtgaccagt ggagccccc ccccgagtca gaccagaaag 720  
 ctaggtogtg ggtcagctct gaggatgtat acccctgggt ggagagggag acctagagat 780  
 ctggctgttg ggcgggcatg gggggtgaag ggccactggg accctcagcc ttgtttgtac 840  
 tgtatgcctt cagcattgcc taggaacacg aagcagatc agtccatcca gagggaccgg 900  
 agttatgaca agcttcccaa atattttgct ttatcagccg atatcaacac ttgtatctgg 960

cctctgtgcc cagcagtgcc ttgtgcaatg tgaatgtacc gtctctgcta aaccaccatt 1020  
 ttatattgggtt ttgtttttgtt tggttttctc ggatacttgc caaaatgaga ctctccgtcg 1080  
 gcagctgggg gaagggtctg agactctctt tccttttgggt tttgggatta cttttgatcc 1140  
 tgggggaacca atgaggtgag gggggttctc ctttgccctc agctttccca gccctccggc 1200  
 ctgggctgcc cacaaggctt ctccccaga ggccctggct cctggtcggg aaggagggtg 1260  
 cctcccgcca acgcatcact ggggctggga gcagggaagg gaattc 1306

<210> 27  
 <211> 216  
 <212> DNA  
 <213> Homo sapiens

<400> 27  
 agcgagagcg cccccgagca gcgcccgcgc cctccgcgc ttctccgcgc ggacctcgag 60  
 cgaaagacgc ccgcccgcgc ccagccctc gcctccctgc ccaccgggca caccgcgcgc 120  
 ccaccccgac ccgctgcgc acggcctgtc cgctgcacac cagcttggtg gcgtcttcgt 180  
 cgccgcgctc gcccgggct actcctgcgc gccaca 216

<210> 28  
 <211> 687  
 <212> DNA  
 <213> Homo sapiens

<400> 28  
 taaatgctac ctgggtttcc agggcacacc tagacaaaca rgggagaaga gtgtcagaat 60  
 cagaatcatg gagaaaatgg gcgggggtgg tgtgggtgat gggactcatt gtagaaagga 120  
 agccttgctc attcttgagg agcattaagg tatttcgaaa ctgccaaggg tgctggtgcg 180  
 gatggacact aatgcagcca cgattggaga atactttgct tcatagtatt ggagcacatg 240  
 ttactgcttc attttgagc ttgtggagtt gatgacttct tgtttctgt ttgtaaatta 300  
 tttgctaagc atattttctc taggcttttt tccttttggg gttctacagt cgtaaaagag 360  
 ataataagat tagttggaca gtttaaagct tttattcgtc ctttgacaaa agtaaatggg 420  
 agggcattcc atcccttctt gaagggggac actccatgag tgtctgtgag aggcagctat 480  
 ctgcactcta aactgcaaac agaaatcagg tgttttaaga ctgaatgttt tatttatcaa 540  
 aatgtagctt ttggggaggg aggggaaatg taatactgga ataatttgta aatgatttta 600  
 attttatatt cagtgaagag attttattta tggaattaac catttaataa agaaatattt 660

acctaaaaaa aaaaaaaaaa aaaaaaa 687

<210> 29  
 <211> 310  
 <212> DNA  
 <213> Homo sapiens

<400> 29  
 cggccccaga aaacccgagc gagtaggggg cggcgcgcag gagggaggag aactgggggc 60  
 gcgggaggct ggtgggtgtc gggggtggag atgtagaaga tgtgacgccg cgccccggcg 120  
 ggtgccagat tagcggacgg ctgccgcggg ttgcaacggg atcccgggcg ctgcagcttg 180  
 ggaggcggct ctccccaggc ggcgtccgcg gagacacca tccgtgaacc ccagggtccc 240  
 ggccgccggc tcgccgcgca ccagggggcg gggacagaa gagcgggcga gcggctcgag 300  
 gctgggggac 310

<210> 30  
 <211> 5882  
 <212> DNA  
 <213> Homo sapiens

<400> 30  
 ctgctaagag ctgattttta tggccacatc taatctcatt tcacatgaaa gaagaagtat 60  
 attttagaaa tttgttaatg agagtaaaag aaaataaatg tgtatagctc agtttggata 120  
 attggtcaaa caatttttta tccagtagta aaatatgtaa ccattgtccc agtaaagaaa 180  
 aataacaaaa gttgtaaaat gtatatcttc ccttttatat tgcatctgct gttaccagct 240  
 gaagcttacc tagagcaatg atctttttca cgcatttgct ttattcgaaa agaggctttt 300  
 aaaatgtgca tgtttagaaa caaaatttct tcatggaaat catatacatt agaaaatcac 360  
 agtcagatgt ttaatcaatc caaaatgtcc actatttctt atgtcattcg ttagtctaca 420  
 tgtttctaaa catataaatg tgaatttaat caattccttt catagtttta taattctctg 480  
 gcagttcctt atgatagagt ttataaaaca gtcctgtgta aactgctgga agttcttcca 540  
 cagtcaggtc aattttgtca aacccttctc tgtaccata cagcagcagc ctagcaactc 600  
 tgctggtgat gggagttgta ttttcagtct tcgccaggtc attgagatcc atccactcac 660  
 atcttaagca ttcttctctg caaaaattta tggatgaatga atatggcttt aggcggcaga 720  
 tgatatacat atctgacttc caaaagctc caggatttgt gtgctgttgc cgaatactca 780  
 ggacggacct gaattctgat tttataccag tctcttcaaa aacttctoga accgctgtgt 840



ctccctacgta aaaaaagaga tgtacaaatc aataataatt acacttttag aaactgtatc 900  
 atcaaagatt ttcagttaaa gtagcattat gtaaaggctc aaaacattac cctaacaaag 960  
 taaagttttc aatacaaatt ctttgcccttg tggatatcaa gaaatoccaa aatattttct 1020  
 taccactgta aattcaagaa gcttttgaaa tgcctgaatat ttctttggct gctacttgga 1080  
 ggcttatcta cctgtacatt ttgggggtca gctcttttta acttcttgct gctctttttc 1140  
 ccaaaaggta aaaatataga ttgaaaagtt aaaacatttt gcatggctgc agttcccttg 1200  
 tttcttgaga taagattcca aagaacttag attcatttct tcaacaccga aatgctggag 1260  
 gtgtttgatc agttttcaag aaacttgga tataaataat ttataattc aacaaagggt 1320  
 ttcacatttt ataagggtga tttttcaatt aaatgcaaatt ttgtgtggca ggatttttat 1380  
 tgccattaac atatttttgt ggctgctttt tctacacatc cagatggctc ctctaactgg 1440  
 gctttctcta attttgtgat gttctgtcat tgtctcccaa agtathtagga agaagccctt 1500  
 taaaaagctg ccttcctcta ccactttgct ggaaagcttc acaattgtca cagacaaaga 1560  
 tttttgttcc aataactggt ttgcctctat ttttcttggt tgtcaaatag taaatgatat 1620  
 ttgcccttgc agtaattcta ctggtgaaaa acatgcaaag aagaggaagt cacagaaaca 1680  
 tgtctcaatt cccatgtgct gtgactgtag actgtcttac catagactgt cttacccatc 1740  
 ccctggatat gctcttggtt ttccctcta atagctatgg aaagatgcat agaaagagta 1800  
 taatgtttta aaacataagg cattcatctg ccatttttca attacatgct gacttccctt 1860  
 acaattgaga tttgccata ggtaaacaat ggtagaaac aactgaaagc ataaaagaaa 1920  
 aatctaggcc ggggtgcagt gctcatgctt atattccctg cactttggga ggccaaagca 1980  
 ggaggatcgc ttgagcccag gagttcaaga ccaacctggg gaaaccccgct ctotacaaaa 2040  
 aaacacaaaa aatagccagg catggtggcg tgtacatgtg gtctcagata cttgggaggc 2100  
 tgagggtgga ggggtgatca cttgaggctg agaggtaag gttgcagtga gccataatcg 2160  
 tgccactgca gtccagccta ggcaacagag tgagactttg tctcaaaaaa agagaaattt 2220  
 tccttaataa gaaaagtaat ttttactctg atgtgcaata cttttgttat taaatttatt 2280  
 atttaagatg gtagcactag tcttaaattg tataaaatat ccctaacaat gtttaaatgt 2340  
 ccatttttat tcattatgct ttgaaaaata attatgggga aatacatgtt tgttattaaa 2400  
 tttattatta aagatagtag cactagtctt aaatttgata taacatctcc taacttggtt 2460  
 aaatgtcoat ttttattctt tatgcttgaa aataaattat ggggatccta tttagctctt 2520

agtaccacta atcaaaagtt cggcatgtag ctcagatct atgctgtttc tatgtcgtgg 2580  
 aagcaccgga tgggggtagt gagcaaactt gccctgctca gcagtcacca tagcagctga 2640  
 ctgaaaatca gcactgcctg agtagttttg atcagtttaa cttgaatcac taactgactg 2700  
 aaaattgaat gggcaaataa gtgcttttgt ctccagagta tgcgggagac ccttccacct 2760  
 caagatggat atttcttccc caaggatttc aagatgaatt gaaattttta atcaagatag 2820  
 tgtgctttat tctgttgtat tttttattat tttaatatac tgtaagccaa actgaaataa 2880  
 catttgctgt tttataggtt tgaagaacat aggaaaaact aagaggtttt gtttttattt 2940  
 ttgctgatga agagatatgt ttaaatatgt tgtattgttt tgtttagtta caggacaata 3000  
 atgaaatgga gtttatattt gttatttcta tttgtttata ttttaataata gaattagatt 3060  
 gaaataaaat ataatgggaa ataactctga gaatgtgggt ttcttggtgt ttctctgac 3120  
 tctagtgcac tgatgatctc tgataaggct cagctgcttt atagttctct ggctaatagca 3180  
 gcagatactc ttcttgccag tggtaatacg attttttaag aaggcagttt gtcaatttta 3240  
 atcttgtgga tacctttata ctcttagggt attattttat acaaaagcct tgaggattgc 3300  
 attctatttt ctatatgacc ctcttgatat ttaaaaaaca ctatggataa caattcttca 3360  
 tttacctagt attatgaaag aatgaaggag ttcaaacaaa tgtgtttccc agttaactag 3420  
 ggtttactgt ttgagccaat ataaatgttt aactgtttgt gatggcagta ttctaaagt 3480  
 acattgcatg ttttcctaaa tacagagttt aaataatttc agtaattctt agatgattca 3540  
 gcttcatcat taagaatata ttttgtttta tgttgagtta gaaatgcctt catatagaca 3600  
 tagtctttca gacctctact gtcagttttc atttctagct gctttcaggg ttttatgaat 3660  
 tttcaggcaa agctttaatt tatactaagc ttaggaagta tggctaatac caacggcagt 3720  
 tttttcttc ttaattccac atgactgagg catatatgat ctctgggtag gtgagttgtt 3780  
 gtgacaacca caagcacttt tttttttttt aaagaaaaaa aggtagtga tttttaatca 3840  
 tctggacttt aagaaggatt ctggagtata cttaggcctg aaattatata tatttggctt 3900  
 ggaaatgtgt ttttcttcaa ttacatctac aagtaagtac agctgaaatt cagaggaccc 3960  
 ataagagttc acatgaaaaa aatcaattca tttgaaaagg caagatgcag gagagaggaa 4020  
 gccttgcaaa cctgcagact gctttttgcc caatatagat tgggtaaggc tgcaaaacat 4080  
 aagcttaatt agctcacatg ctctgctctc acgtggcacc agtggatagt gtgagagaat 4140  
 taggctgtag aacaaatggc ctctcttttc agcattcaca ccactacaaa atcatctttt 4200  
 atatcaacag aagaataagc ataaactaag caaaaggcca ataagtaact gaaaccaaga 4260

ttggctagag atatatctta atgcaatcca ttttctgatg gattggttacg agttggctat	4320
ataatgtatg tatgggtatgt tgatttgtgt aaaagtttta aaaatcaagc ttttaagtaca	4380
tggacattttt taaataaaat atttaaagac aatttagaaa attgccttaa tatcattggt	4440
ggctaaatag aataggggac atgcatatta aggaaaaggt catggagaaa taatattggt	4500
atcaaacaaa tacattgatt tgtcatgata cacattgaat ttgatccaat agtttaagga	4560
ataggtagga aaatttggtt tctatttttc gatttctctgt aaatcagtga cataaataat	4620
tcttagctta ttttatatgt ccttgtctta aatactgagc tcagtaagtt gtgttagggg	4680
attattttctc agttgagact ttcttatatg acattttact atgttttgac ttcctgacta	4740
ttaaaaataa atagtagaaa caattttcat aaagtgaaga attatataat cactgcttta	4800
taactgactt tatttatatgt atttcaaagt tcatttaaag gctactattc atcctctgtg	4860
atggaatggt caggaatttg ttttctcata gtttaattcc aacaacaata ttagtcgtat	4920
ccaaaataac ctttaatgct aaactttact gatgtatata caaagcttct ctttttcaga	4980
cagattaatc cagaagcagt cataaacaga agaatagggtg gtatgttcct aatgatatta	5040
tttctactaa tggaataaac tgtaatatta gaaattatgc tgctaattat atcagctctg	5100
aggtaatttc tgaaatgttc agactcagtc ggaacaaatt ggaaaattta aatttttatt	5160
cttagctata aagcaagaaa gtaaacacat taatttctctc aacattttta agccaattaa	5220
aaatataaaa gatacacacc aatatcttct tcaggctctg acaggcctcc tggaaaacttc	5280
cacataatgtt tcaactgcag tataaagtca gaaaataaag ttaacataac tttcactaac	5340
acacacatat gtagatttca caaaatccac ctataattgg tcaaagtggt tgagaatata	5400
tttttttagta attgcatgca aaatttttct agcttccatc ctttctccct cgtttcttct	5460
ttttttgggg gagctggtaa ctgatgaaat cttttccac ctttctctt caggaaatat	5520
aagtggtttt gtttggttaa cgtgatacat tctgtatgaa tgaaacattg gagggaaaca	5580
tctactgaat ttctgtaatt taaaatattt tgctgctagt taactatgaa cagatagaag	5640
aatcttacag atgctgctat aaataagtag aaaatataaa tttcatcact aaaatatgct	5700
attttaaaat ctatttctta tattgtatgt ctaatcagat gtattactct tattatttct	5760
attgtatgtg ttaatgattt tatgtaaaaa tgtaattgct tttcatgagt agtatgaata	5820
aaattgatta gtttgtgttt tcttgtctcc cgaaaaaaaa aaaaaaaaaa aaaaaaaaaa	5880
aa	5882

<210> 31  
 <211> 310  
 <212> DNA  
 <213> Homo sapiens

<400> 31  
 cggccccaga aaacccgagc gagtaggggg cggcgcgag gagggaggag aactgggggc 60  
 gcgggaggct ggtgggtgtc gggggtggag atgtagaaga tgtgacgccg cggccccgcg 120  
 ggtgccagat tagcggacgg ctgcccggg ttgcaacggg atcccgggcg ctgcagcttg 180  
 ggaggcggct ctccccaggc ggcgtccgcg gagacacca tccgtgaacc ccagggtccc 240  
 ggccgccggc tcgccgcgca ccaggggccg gcggacagaa gagcggccga gcggctcgag 300  
 gctgggggac 310

<210> 32  
 <211> 3212  
 <212> DNA  
 <213> Homo sapiens

<400> 32  
 tgagggcgcc aggcaggcgg gcgccaccgc caccgcagc gagggcggag ccggccccag 60  
 gtgctccoct gacagtccct cctctccgga gcattttgat accagaaggg aaagcttcat 120  
 tctccttggt gttgggtgtt ttttcctttg ctctttcccc ctccatctc tgacttaagc 180  
 aaaagaaaaa gattacccaa aaactgtctt taaaagagag agagagaaaa aaaaaatagt 240  
 atttgcataa ccctgagcgg tgggggagga gggttgtgct acagatgata gaggatttta 300  
 taccocaata atcaactcgt ttttatatta atgtacttgt ttctctgttg taagaatagg 360  
 cattaacaca aaggaggcgt ctggggagag gattaggttc catcctttac gtgtttaaaa 420  
 aaaagcataa aaacatttta aaaacataga aaaattcagc aaaccatttt taaagtagaa 480  
 gagggtttta gtagaaaaa catattcttg tgcttttccct gataaagcac agctgtagtg 540  
 gggttctagg catctctgta ctttgcttgc tcatatgcat gtagtcactt tataagtcac 600  
 tgtatgttat tatattccgt aggtagatgt gtaacctctt caccttattc atggctgaag 660  
 tcacctcttg gttacagtag cgtagcgtgg ccgtgtgcat gtcctttgcg cctgtgacca 720  
 ccacccaac aaaccatcca gtgacaaaac atccagtga ggtttgtcgg gcaccagcca 780  
 gcgtagcagg gtccgggaaag gccacctgtc ccactcctac gatacgctac tataaagaga 840  
 agacgaaata gtgacataat atattctatt tttatactct tctatttttt gtagtgacct 900  
 gtttatgaga tgctggtttt ctacccaacg gccctgcagc cagctcacgt ccaggttcaa 960

cccacagcta cttggtttgt gttcttcttc atattctaaa accattccat ttccaagcac 1020  
 tttcagtcca ataggtgtag gaaatagcgc tgtttttgtt gtgtgtgcag ggagggcagt 1080  
 tttctaattgg aatggtttgg gaatatccat gtacttggtt gcaagcagga ctttgaggca 1140  
 agtgtgggcc actgtgggtg cagtggaggt ggggtgtttg ggaggtgcg tgccagtcaa 1200  
 gaagaaaaag gtttgcattc tcacattgcc aggatgataa gttcctttcc ttttctttaa 1260  
 agaagttgaa gtttaggaat cctttggtgc caactggtgt ttgaaagtag ggacctcaga 1320  
 ggtttaccta gagaacaggt ggtttttaag ggttatctta gatgtttcac accggaaggt 1380  
 ttttaaacac taaaatatat aatttatagt taaggctaaa aagtatattt attgcagagg 1440  
 atgttcataa ggccagtatg atttataaat gcaatctccc cttgatttaa acacacagat 1500  
 acacacacac acacacacac acacacaaac cttctgcctt tgatgttaca gatttaatac 1560  
 agtttatattt taaagataga tccttttata ggtgagaaaa aaacaatctg gaagaaaaaa 1620  
 accacacaaa gacattgatt cagcctgttt ggcgtttccc agagtcatct gattggacag 1680  
 gcatgggtgc aaggaaaatt aggttactca acctaagttc ggttccgatg aattcttato 1740  
 ccctgcccct tccttttaaaa aacttagtga caaaatagac aatttgcaca tcttggtat 1800  
 gtaattcttg taatttttat ttaggaagtg ttgaaggag gtggcaagag tgtggaggct 1860  
 gacgtgtgag ggaggacagg cgggaggagg tgtgaggagg aggcctccga ggggaagggg 1920  
 cggtgcccac accggggaca ggccgcagct ccattttctt attgcgctgc taccgttgac 1980  
 ttccaggcac ggtttggaaa tattcacatc gcttctgtgt atctctttca cattgtttgc 2040  
 tgctattgga ggatcagttt tttgttttac aatgtcatat actgccatgt actagtttta 2100  
 gttttctctt agaacattgt attacagatg ccttttttgt agttttttt ttttttatgt 2160  
 gatcaatttt gacttaattg gattactgct ctattccaaa aaggttgctg tttcacaata 2220  
 cctcatgctt cacttagcca tgggtggacc agcgggcagg ttctgcctgc tttggcgggc 2280  
 agacacgcgg gcgcgatccc acacaggctg gcgggggcgg gccccgagge cgcgtgcgtg 2340  
 agaaccgcgc cgggtgtccc agagaccagg ctgtgtccct cttctcttcc ctgcgcctgt 2400  
 gatgtggggc acttcactgt atcgggggcg tagcatcata gtagttttta cagctgtgtt 2460  
 attctttgcg tgtagctatg gaagttgcat aattattatt attattatta taacaagtgt 2520  
 gtcttaoctg ccaccacggc gttgtacctg taggactctc attcgggatg attggaatag 2580  
 cttctggaat ttgttcaagt tttgggtatg tttaatctgt tatgtactag tgttctgttt 2640

```

gttattgttt tgtaattac accataatgc taatttaaag agactccaaa tctcaatgaa 2700
gccagctcac agtgetgtgt gccccggtca cctagcaagc tgccgaacca aaagaatttg 2760
caccocgctg cgggcccacg tggttggggc cctgccctgg cagggtcacc ctgtgctcgg 2820
aggccatctc gggcacaggc ccaccccgcc ccacccctcc agaacacggc tcacgcttac 2880
ctcaaccatc ctggctgctg cgtctgtctg aaccacgctg gggccttgag ggacgctttg 2940
tctgtcgtga tggggcaagg gcacaagtc tggatgttgt gtgtatcgag aggccaaagg 3000
ctgggtggca gtgcacgggg cacagcggag tctgtcctgt gacgcgcaag tctgagggtc 3060
tgggcgcgcg gcggtgggt ctgtgcattt ctggttgcac cgcggcgctt ccagcacca 3120
acatgtaacc ggcattgttc cagcagaaga caaaaagaca aacatgaaag tctagaaata 3180
aaactggtaa aacccccaaa aaaaaaaaaa aa 3212

```

```

<210> 33
<211> 1043
<212> DNA
<213> Homo sapiens

```

```

<220>
<221> misc_feature
<222> (409)..(444)
<223> n = a, t, g or c

```

```

<400> 33
gcacgcggc gagcttggt gcttctggg cctgtgtgg cctgtgtgc ggaaagatgg 60
agcaagaagc cgagcccag gggcgccgc gacccctctg accgagatcc tgctgctttc 120
gcagccagga gcacgtccc tccccgatt agtgcgtacg agcgcccagt gccctggccc 180
ggagagtgga atgatccccg agggccaggg cgtcgtgctt ccgcgcgcc cgtgaaggaa 240
actggggagt cttgaggga ccccgactcc aagcgcgaaa accccggatg gtgaggagca 300
ggtactggcc cggcagcag cggtcacttt tgggtctgg ctctgacgg gtccctcta 360
tcgctggttc ccagcctctg ccggttcga gcctttgtgc ggttcgtgnc tgggggctcg 420
gggcgcgggg cgcggggcat gggncacgtg gctttgcgga ggttttggtg gactggggct 480
agacagtccc cgccaggag gagggcggga ttctcgacgg ctctcgcggc ggtgggggtg 540
ggggtggttc ggaggtctcc gcgggagttc agggtaaagg tcacggggcc ggggctcgg 600
gccgcttcgg cgcgggaggt ccggtatgac gcagtgcctg tcgggtcact agtgtgaacg 660

```

ctgcgcgtag tctggggcggg attggggccgg ttcagtgggc aggttgactc agcttttcct 720  
 cttgagctgg tcaagttcag acacgttcog aaactgcagt aaaaggagtt aagtcctgac 780  
 ttgtctccag ctggggctat ttaaaccatg ctttttccca gctgtgttca gtggcgattg 840  
 gagggtagac ctgtgggcac ggacgcacgc cactttttct ctgctgatcc aggtaagcac 900  
 cgacttgctt gtagcttttag ttttaactgt tgtttatgtt ctttatatat gatgtatttt 960  
 ccacagatgt ttcagtattt ccagttttca tctgtctttt tttttccttg taggcaaatg 1020  
 tgcaatacca acatgtctgt acc 1043

<210> 34  
 <211> 1153  
 <212> DNA  
 <213> Homo sapiens

<400> 34  
 tagttgacct gtctataaga gaattatata tttctaacta tataacccta ggaatttaga 60  
 caacctgaaa tttattcaca tatatcaaag tgagaaaatg cctcaattca catagatttc 120  
 ttctcttttag tataattgac ctactttggt agtggaaatag tgaatactta ctataatttg 180  
 acttgaatat gtagctcatc ctttacacca actoctaatt ttaaataatt tctactctgt 240  
 cttaaattgag aagtacttgg tttttttttt cttaaattatg tatatgacat ttaaattgaa 300  
 cttattattt tttttgagac cgagtcttgc tctgttaccc aggctggagt gcagtgggtg 360  
 atcttggctc actgcaagct ctgccctccc cgggttcgca ccattctcct gcctcagcct 420  
 cccaattagc ttggcctaca gtcactctgc accacacctg gctaattttt tgtactttta 480  
 gtagagacag ggtttcaccg tgtagccag gatggtctcg atctcctgac ctctgatcc 540  
 gccacctcg gcctcccaaa gtgtgggat tacaggcatg agccaccgtg ctctccagcc 600  
 taggcaacag agtgagactc tgtctccaaa aaaaaaaaaa aaaaaggagg actataaacac 660  
 ccccagggaaggaggacaggt gggacattct tattcttaat ttaaataaat tgacagggga 720  
 aagttgggcc actcttgagc ttgtgggtgc tcaccaggtt gaccccaaaa aaagaagcct 780  
 tccacaaaac attaatattt ttccctaata taccgcctc tgtgagttaa gggataatgc 840  
 atcaggactc ttgcaaccag acaaaaattt ttaaaaacgc caottggggg ggaggcgggt 900  
 cctccttggg gattcgcctt tgtgggagag aaaactgcac agacttgggc aaataatgtt 960  
 ttttgcacc ccaaaacgta ttcgogagac atttcattag aacgaagctt taccctaata 1020  
 ttgaactccc catttaaaca gtttcacac acacttaggg agatttttcc ctctgtgagt 1080

tccgcagaac aatagttgga cgggaataga accctgaaac actttagttc accacgaact 1140  
attatagggc ggg 1153

<210> 35  
<211> 334  
<212> DNA  
<213> Homo sapiens

<400> 35  
tgactatcca gctctgagag acgggagttt ggagttgccc gctttacttt ggttgggttg 60  
ggggggggcgg cgggctgttt tgttcctttt cttttttaag agttgggttt tcttttttaa 120  
ttatccaaac agtgggcagc ttctccccc acaccaagt atttgcacaa tatttgtgcg 180  
gggtatgggg gtgggttttt aaatctcgtt tctcttggaac aagcacaggg atctcgttct 240  
cctcattttt tgggggtgtg tggggacttc tcaggtcgtg tcccagcct tctctgcagt 300  
cccttctgcc ctgccgggcc cgtcgggagg cgcc 334

<210> 36  
<211> 543  
<212> DNA  
<213> Homo sapiens

<400> 36  
tagctcagga ccttggtctg gcctggctgt catgtaggtc aggaccttg ctggacctgg 60  
aggccctgcc cagccctgct ctgcccagcc cagcaggggc tccaggcctt ggctggcccc 120  
acatcgccctt ttctccccc acacctcgtt gcacttgtgt ccgaggagcg aggagcccct 180  
cgggcccttg gtggcctctg ggccctttct cctgtctccg ccactccctc tggcggcgct 240  
ggccgtggct ctgtctctct gaggtgggtc gggcgccctc tgcccgcccc ctcccacacc 300  
agccaggctg gtctcctcta gcctgtttgt tgtgggtgtg gggatatatt tgaaccact 360  
gggccccag cccctctttt gcgaccctt gtctgacct gttctcggca ccttaaatta 420  
ttagaccccg gggcagtcag gtgctccgga caccgaagg caataaaaca ggagccgtga 480  
aaaaaaaaa aaaaaaaaaa aaaaaaaaaa aaaaaaaaaa aaaaaaaaaa aaaaaaaaaa 540  
aaa 543

<210> 37  
<211> 511  
<212> DNA  
<213> Homo sapiens

<400> 37



gctcagcaag ggggccgtcc ttctctgtca ctgtctcttt tgccctgttg aattctgtct 60  
 gcctctctgg gactctgcct gtctcactct ttctgtctgt gcctctctc actcttggtc 120  
 tttctgcctg aatcacagcc ctccagtttt ctgtcctcat gcatttgctt ttgtggctct 180  
 ttccgtcttt ctgcccttga caccatcccc tctccagtg cttccctct gcttccagat 240  
 cgcttcatga cttaggcagg gaaacagagg tcagggcctc cttccaggct tccctctgca 300  
 tcttactgag tatgcaggtc ggaagagcct cgggtcctgc ctccgcgggt ggccatagagc 360  
 caaaggaagg cggagcccgt cggggcgagg ttggccctta gggccacctc ataaagcctg 420  
 gggcgagggg cacaacggcc ttgggaagga gccctgttgg ggccgtccag tccccagac 480  
 ctccagcgt cagtcgcgga tctgcagtgt c 511

<210> 38  
 <211> 458  
 <212> DNA  
 <213> Homo sapiens

<400> 38  
 tagtagggac cagtgaccat cacatccctt caagagtcct gaagatcaag ccagttctcc 60  
 ttccctgcag agctttggcc attaccacct gacctcttgc tgccagctaa taagaagtgc 120  
 caagtggaca gtctggccac tgtcaaggca ggaaggggc catgactttt ctgccctgcc 180  
 ctccagcctg tgccctgcct cccaaacccc attagtctag ccttgtagct gttactgcaa 240  
 gtgtttcttc tggttagtgc tgttttctaa agccaggact attccctttc ctccccagga 300  
 atatgtgttt tcttttgtct taatcgatct ggtaggggag aaatggcgaa tgcatacac 360  
 atgagatggg atatccttgc gatgtacaga atcagaaggg ggtttgacag catcataaac 420  
 aggctgactg gcaggaatga aaaaaaaaaa aaaaaaaaaa 458

<210> 39  
 <211> 270  
 <212> DNA  
 <213> Homo sapiens

<400> 39  
 ggggccgccc agagccgcag cgcgcctcgc ccgcgcgcc ccccccgcg gcccgcgccc 60  
 gcgaattgcg cccgcgcgcc tccctctcgc ccccgagac aaagaggaga gaaagtttgc 120  
 gcggccgagc gggcaggtga ggagggtgag ccgcgcggag gggccgcct cggcccgccc 180  
 tcagcccccg cccgcgcgcc cagcccgccc ccgcgcgag cgcgcggacc ccccgagccc 240  
 ggcccgcgcc gccagcccc ccggcccgcc 270

<210> 40  
 <211> 751  
 <212> DNA  
 <213> Homo sapiens

<220>  
 <221> misc\_feature  
 <222> (535)..(739)  
 <223> n = a, t, g or c

<400> 40  
 taagcaggcc tccaacgccc ctgtggccaa ctgcaaaaaa agcctccaag ggtttcgact 60  
 ggtccagctc tgacatccct tcctggaaac agcatgaata aaacactcat cccatgggtc 120  
 caaattaata tgattctgct ccccccttct ccttttagac atggttgttg gtctggaggg 180  
 agacgtgggt ccaaggtcct catcccatcc tccctctgcc aggcactatg tgtctggggc 240  
 ttgatccctt ggggtgcaggc agggctggga cagcggtctt cctcccagt cctgccttg 300  
 gcaccgtcac agatgccaaag caggcagcac ttagggatct cccagctggg ttagggcagg 360  
 gcctggaaat gtgcattttg cagaaacttt tgagggtcgt tgcaagactg ttagcaggc 420  
 ctaccaggtc cctttcatct tgagagggac atggcccctt gttttctgca gcttccacgc 480  
 ctctgcactc cctgcccctg gcaagtgtct ccatcgcccc cgggtgccac catgnagctc 540  
 cccgcacctg actcccccca catccaaggc cagccctgga accagtgggc tagttccttg 600  
 aaggaagccc cactcattcc tattaatccc tcagaattcc cgggggggagc ctccctcct 660  
 gaaccttggt aaaaaatggg gaacgagaaa aacccccgct tggagctgtg cgtttccagc 720  
 ccctacttga gagncttttt tttgggggcc g 751

<210> 41  
 <211> 229  
 <212> DNA  
 <213> Homo sapiens

<400> 41  
 cgcgcggggc ccggctcggc ccgaccggc tccgcggggc caggcggggc ccagcgcact 60  
 cggagcccga gcccgagccg cagccgccgc ctggggcgct tgggtcggcc tcgaggacac 120  
 cggagagggg cgccacgccg ccgtggccgc agatttgaaa gaagccgaca ctaaaccacc 180  
 aatatacaac aaggccattt tgtcaaacga gagtcagcct ttaacgaaa 229

<210> 42

<211> 233  
 <212> DNA  
 <213> Homo sapiens

<400> 42  
 tagcagagag tcctgagcca ctgccaacat ttcccttctt ccagttgcac tattctgagg 60  
 gaaaatctga cacctaagaa atttactgtg aaaaagcatt ttaaaaagaa aagggttttag 120  
 aatatgatct attttatgca tattgtttat aaagacacat ttacaattta cttttaatat 180  
 taaaaattac catattatga aaaaaaaaaa aaaaaaaaaa aaaaaaaaaa aaa 233

<210> 43  
 <211> 349  
 <212> DNA  
 <213> Homo sapiens

<400> 43  
 ggcacgaggg gcgagaggaa gcagggagga gagtgatttg agtagaaaag aaacacagca 60  
 ttccaggctg gcccacctc tatattgata agtagccaat gggagcgggt agccctgac 120  
 cctggccaat ggaaactgag gtaggcgggt catcgcgctg gggctctgtag tctgagcgt 180  
 acccggttgc tgctgcccaa ggaccgcgga gtcggacgca ggcagaccat gtggaccctg 240  
 gtgagctggg tggccttaac agcagggctg gtggctggaa cgcggtgcc agatggtcag 300  
 ttctgccttg tggcctgctg cctggacccc ggaggagcca gctacagct 349

<210> 44  
 <211> 337  
 <212> DNA  
 <213> Homo sapiens

<400> 44  
 tgagggacag tactgaagac tctgcagccc tcgggacccc actcggaggg tgccctctgc 60  
 tcaggcctcc ctagcacctc cccctaacca aattctccct ggacccatt ctgagctccc 120  
 catcaccatg ggaggtgggg cctcaatcta aggccttccc tgtcagaagg gggttgtggc 180  
 aaaagccaca ttacaagctg ccatccctc cccgtttcag tggaccctgt ggccaggtgc 240  
 tttccctat ccacaggggt gtttgtgtgt gtgcgcgtgt gcgtttcaat aaagtttgta 300  
 cactttcaaa aaaaaaaaaa aaaaaaaaaa aaaaaaa 337

<210> 45  
 <211> 1700  
 <212> DNA  
 <213> Homo sapiens

```

<400> 45
tgtttgcat aagttcatag attataatgt gtaatggaat caacacccaaa tgcaaattag 60
aaagagagcc cactttgctc acccagtcac gtcttcccat gtaaccatag aacgttgggg 120
tcctgtgtct ttctagatcc acagtcttgc tctcagaaca ggctagccac accacaggcc 180
tagtgccagg acccatggcc tttttttaag ctcagactcc cttctgtgaa cagcaatata 240
cccacaactt gtacaacatt ggtgcttcc t gcaagggtca cagaactatt tgatacgaaa 300
atgttcattg acttacacac aagagaagca caaaataaaa aattaataat taatttaattg 360
tctttgaaaa tgtaccatgt atttttacat ttgggggtcat aagaattgta ttacacttaa 420
gaatgcaata caatttgaag atcagatttt tctccctttg tgagaatttc tcagtatgtg 480
tgatgactac caagaaatca tagccagtca taaattcagt gagttactca taaacgaaca 540
agaaccacct acttcttggg gaggtaggtc tgcttccctt caactcagga tacaactgct 600
ttcaactgct ttcttcacat tagctgacta attagctaga agcctgtcgt aaacaatttt 660
atggttgact ccttccctgg gctcagggtt ccttagaaca gagagggtccc caaatcccg 720
tctgtggcct gtccgcctaa gctctgcctc ctgccagatc agcaggcagc attagattct 780
cataggagct ggacgcctat tgtgaactgc gcatgtgcgg gatccagatt gtgcactctt 840
tatgagaatc taactaatgc ttgatgatct atctgaacca gaacaatttc atcctgaaac 900
catccccac caatccatag aaatactgtc ttccacaaaa atgatccctg gtgccaaaaa 960
tgtagagac cactccctca aaactctctt cttagctctc acctcctgta ttactatctc 1020
atctcagtac attgaagccc ccatcttttc cccatggatg cctcatttcc tattagggag 1080
gcattttttt attttttgtt tttatttttt tccgagacgg agtctcgtc tgtagccaag 1140
gctggagtgc agtggcgcca tctcggtcca ctgcaagctc cgcctcccg gttcacgcca 1200
ttctcctgcc tcagcctccc aagtagctgg gactacaggc gccgcacta cgcgcggcta 1260
attttttgta ttttttagtag agacgggggt tcaccgtggt agccaggatg gtctcgatct 1320
cctgacctcg tgatccgccc gccttggcct cccaaagtgc tgggattaca ggcgtgagac 1380
cgcgcccggc cgtcatttgg tatgtotaa tgtgcctcag gacctagcac agtccctggt 1440
accagtaga gacctatgta atgttcgtta ttcaataata aatacatgaa ttaaagagtg 1500
agagtggatt ttgtaatgtt acgactgata gagaaatact cagtgattct aagggatggg 1560
gaagaacggt tggagctaga ggttgtgctc aggaactat taaatagacg ttccgcagga 1620

```

agggattgac gaagtgtgag gttaatgagg aagggaat agaataaaa atttggtggt 1680  
 ggaaaagatc tgattcatga 1700

<210> 46  
 <211> 2419  
 <212> DNA  
 <213> Homo sapiens

<400> 46  
 taaccagcgg gcccttggtc aagtgtggtc tctgtgtgcc ttgccttcca tttccctct 60  
 gcaccagaa cagtgggtgc aacattcatt gccaggggcc caaagaaaga gctacctgga 120  
 ccttttggtt tctgtttgac aacatgttta ataaataaaa atgtcttgat atcagtaaga 180  
 atcagagtct tctcactgat tctgggcata ttgatctttc cccattttc tctacttggc 240  
 tgctccctga gaggactgca taggatagaa atgccttttt cttttctttt cgtttttttt 300  
 tttttttttt ttgagatgg agtctcactc tgtcgccag gcttaagtgc aatggcacia 360  
 tctcggtcca ctgcaacctc tctctcctgg gttcaagtga ttctcctgcc tcagcctccc 420  
 aaatagctga gattacaggc atgcaccacc acacctgggt aatttttgtg ttttagtag 480  
 agacagggtt tcaccgtttt ggccagggtg gtcttgaact cctgacctcg ggagatccgc 540  
 ccaccttggc ctctctttgt gctgggatta caggcatgag cactgagcc gggccacttt 600  
 ttcttatca gtcagttttt acaagtcatt agggaggtag actttacctc tctgtgaagg 660  
 aaagtatggt atgttgatct acagagagag atggaaaaat tcagggtc gtagctacta 720  
 agcagaattt ocaagatagg caaattgttt tttctgtcaa ataataagct aatattactt 780  
 ctacaaatat gagaccttgg agagaagttt ccaaggacca agtaccaaca taccaacaga 840  
 ttattatagt ttctctcact cttacacaca cacacacaca tatacacata tgtaatccag 900  
 catgaatacc aaaattcatt cagggtagcc accttttgtc ttaatcgaga gataattttg 960  
 atgtttgaat ggaatgtccc caggatattc tctgtcatg gttattttat ataaaattca 1020  
 aaaaccaatt acattatttc ctctgtaatc ttttacttta tcaactaatg tctggcaagt 1080  
 gtgatgtttt ggggaagtta tagaagattc cgccaggcg cttatctcac gcttgaatc 1140  
 cagcactttg ggaagctgag gcggacagat cagagggtca agagatcaag accatcctgg 1200  
 acaacatggt gaaaccttgt ctctactaaa aatgtgaaa ttagctgggc gtggtggcac 1260  
 acacctatag tccagctac tcgggagggt gaggcaggag aatcgcttga acctaggagg 1320  
 cgagggttgc actgagccga gatcacgcca ctgcactcca gcctgggcga cagagcgaga 1380

ctccatctca aaaaaaaaaa aaaaagaaag atcccagttt atcccagttt atcccttatt 1440  
 ctccctcaat totcaagatt tgtttttaag ttaacataac ttaggttaac acactctttg 1500  
 taaaatacac tgttcaatct acagactcag tggtttagctt cctgttaact aatttctggt 1560  
 gacaggtaact tggatatttt atttagaaag tggttgcaa taaattagtt ataagtcgcc 1620  
 agtttcaactg ccttgtgaac acataattat tgtggtctca gtattcccta tgggtggcttc 1680  
 tctgtctcct ggtattgccc tgaaatgggc caaaagccgt ggctcccaa tgctcagggt 1740  
 atagaacatt gtccaggtaac cacctaggag agcccagcct cactgaaagt attcaaattt 1800  
 aggaatgggt ttgagaagta ggtagctggt atgtgcttag cacaagaatc tctcttcctt 1860  
 gggtttagtct gtttcaaaac tgaaaacact gtcattcctt aagaaaatag gaaaaagtat 1920  
 tccaaacctc tgtcactaga aaatttgcca tattaccaa tctcaaaaac ctctcaggaa 1980  
 atgagaaaagt ccagtttctt ggtaaaactat ttgggcccctt ttctcaagtt ctcttccag 2040  
 tgctatttcc ttgaggtgag gcaaagttac tcaagatcat cgctgccact caaggccttg 2100  
 atagggcaag tgaaaggcat ggaccattat tatattgac acagcataag ctgtgaaaac 2160  
 ccacatcttc tccaaacatc tgcttgagac attatcatcg catagtttgc tctggtgttc 2220  
 agggaaatcg ctgtttcata ggaaatcaca tggcagtggt atgggagtggt ttcttgacct 2280  
 gccgatggta ctggcacctg agcaagcatt cctagtcctt ttggtctgg gcctcttggt 2340  
 ctatcacaac cacaagctgt ttaaaataaa aacgtcaagt cacaggcagg tcattttatc 2400  
 ctgcgtgaat caattgaag 2419

<210> 47  
 <211> 297  
 <212> DNA  
 <213> Homo sapiens

<400> 47  
 tcctcagtgc acagtgtgc ctggtctgag gggacaggag gatcaccctc ttctgtgctt 60  
 cggccagtgt gtcgggctgg gccctgacaa gccacctgag gagaggctcg gagccggggc 120  
 cggaccccg cgtttgcgc ccgtttctct ctagtctcac gaggggtttc ccgctctgca 180  
 ccccccacctc tggacttgcc ttctcttctc ttctccgcgt gtggaggagg ccagcgctta 240  
 ggccggagcg agcctggggg ccgcccgcgc tgaagacatc gcggggaccg attcacc 297

<210> 48  
 <211> 1192

&lt;212&gt; DNA

&lt;213&gt; Homo sapiens

&lt;400&gt; 48

```

tgagcttttt ctttaatttca ttcctttttt tggacactgg tggctcacta cctaaagcag      60
tctatattata tttttotacat ctaatttttag aagcctgggt acaatactgc acaaacttgg      120
ttagttcaat ttttgatccc ctttctactt aatttacatt aatgctcttt tttagtatgt      180
tctttaatgc tggatcacag acagctcatt ttctcagttt tttggtatth aaaccattgc      240
attgcagtag catcatttta aaaaatgcac ctttttattt atttattttt ggctagggag      300
tttatccctt tttcgaatta tttttaagaa gatgccata taatttttgt aagaaggcag      360
taacctttca tcatgatcat aggcagttga aaaattttta cacctttttt ttcacatttt      420
acataaataa taatgctttg ccagcagtag gtggttagcca caattgcaca atatatatttc      480
ttaaaaaata ccagcagtta ctcatggaat atattctgcg tttataaaac tagttttttaa      540
gaagaaatth tttttggcct atgaaattgt taaacctgga acatgacatt gttaatcata      600
taataatgat tcttaaatgc tgtatggtht attatttaaa tgggtaaagc catthacata      660
atatagaaaag atatgcatah atctagaagg tatgtggcat ttatttggat aaaattctca      720
attcagagaa atcatctgat gtttctatag tcactttgcc agctcaaaag aaaacaatac      780
cctatgtagt tgtggaagth tatgctaata ttgtgtaact gatattaaac ctaaatgttc      840
tgcctaccct gttggtataa agatatthtg agcagactgt aaacaagaaa aaaaaaatca      900
tgcattotta gcaaaattgc ctagtatgth aatttgcctca aaatacaatg tttgattthta      960
tgcactthgt cgctattaac atcctthttt tcatgtagat ttcaataatt gagtaattth      1020
agaagcatta ttttaggaat atatagthgt cacagtaaat atcttgthtt ttctatgtac      1080
attgtacaaa tttttcatth cttttgctct ttgtggttgg atctaacact aactgtattg      1140
ttttgttaca tcaataaac atcttctgtg gaccaggaaa aaaaaaaaaa aa      1192

```

&lt;210&gt; 49

&lt;211&gt; 197

&lt;212&gt; DNA

&lt;213&gt; Homo sapiens

&lt;400&gt; 49

```

agacagcctt aaccacggg cgcgggcgag tcgtatgggc aggggcaggc gggagcgacg      60
tggggcgacg ctacgaacg atcagagctg cgggcgacgc aacgaagccc ggaggccgca      120
ggctgcgcgc tcctcgcag cagccgggcg ggcaaaagcc cccagtcctc ggcccccgcg      180

```

caagcgacgc cgggaaa

197

&lt;210&gt; 50

&lt;211&gt; 3293

&lt;212&gt; DNA

&lt;213&gt; Homo sapiens

&lt;400&gt; 50

taattattta tattgtaaag aattttaaca gtcctgggga cttccttgaa ggatcatttt	60
cacttttgct cagaagaaag ctctggatct atcaaataaa gaagtccttc gtgtgggcta	120
catatataga tgttttcatg aagaggagtg aaaagccaga aggatataga caaatgaggc	180
ctaagacctt tcttgccagt aactatactg tcagtagccg gcaaatgtta caagaaatc	240
gggaatccct taggaattta tctaaacat ctgatgctgc taaggctgag cataacatga	300
gtaaaatgtc aaccgaagat cctcgacaag tcagaaatcc acccaaattt gggacgcac	360
ataaagcctt gcaggaaatt cgaaactctc tgcttccatt tgcaaatgaa acaaattctt	420
ctcggagtac ttcagaagtt aatccacaaa tgcttcaaga cttgcaagct gctggatttg	480
atgaggatat ggttatacaa gctcttcaga aaactaaca cagaagtata gaagcagcaa	540
ttgaattcat tagtaaaatg agttaccaag atcctcgacg agagcagatg gctgcagcag	600
ctgccagacc tattaatgcc agcatgaaac cagggaatgt gcagcaatca gttaaccgca	660
aacagagctg gaaaggttct aaagaatcct tagttcctca gaggcattggc ccgccactag	720
gagaaagtgt ggcctatcat tctgagagtc ccaactcaca gacagatgta ggaagacctt	780
tgtctggatc tggatatca gcatttgctc aagctcacc tagcaacgga cagagagtga	840
acccccacc accacctcaa gtaaggagtg ttactcctcc accacctcca agaggccaga	900
ctcccctcc aagaggtaca actccacctc ccccttcatt ggaaccaaac tctcaaaca	960
agcgtatttc tggaacatg gaatacgtaa tctccgaat ctctcctgtc ccacctggg	1020
catggcaaga gggctatcct ccaccacctc tcaacacttc cccatgaat cctcctaata	1080
aaggacagag aggcattagt tctgttctg ttggcagaca accaatcatc atgcagagtt	1140
ctagcaaatt taactttcca tcaggagac ctggaatgca gaatggtact ggacaaactg	1200
atttcatgat acacaaaat gttgtccctg ctggcactgt gaatcggcag ccaccacctc	1260
catatcctct gacagcagct aatggacaaa gcccttctgc tttacaaaca gggggatctg	1320
ctgctccttc gtcatataga aatggaagta ttctcagtc tatgatggtg ccaaacagaa	1380
atagtcataa catggaacta tataacatta gtgtacctgg actgcaaaca aattggcctc	1440



agtcatcttc tgetccagcc cagtcatccc cgagcagtg gcatgaaatc cctacatggc 1500  
 aacctaacat accagtgagg tcaaattctt ttaataaccc attaggaaat agagcaagtc 1560  
 actctgctaa ttctcagcct tctgctacaa cagtcactgc aattacacca gctcctattc 1620  
 aacagcctgt gaaaagtatg cgtgtattaa aaccagagct acagactgct ttagcaccta 1680  
 cacacccttc ttggatacca cagccaattc aaactgttca acccagtcct ttctctgagg 1740  
 gaaccgcttc aaatgtgact gtgatgccac ctggttgctga agctocaaac tatcaaggac 1800  
 caccaccacc ctacccaaaa catctgctgc accaaaaccc atctgttctt ccatacgagt 1860  
 caatcagtaa gcctagcaaa gaggatcagc caagcttgcc caaggaagat gagagtgaat 1920  
 agagttatga aaatgttgat agtggggata aagaaaagaa acagattaca acttcaccta 1980  
 ttactgttag gaaaaacaag aaagatgaag agcgaaggga atctcgtatt caaagttatt 2040  
 ctctcaagc atttaaatc tttatggagc aacatgtaga aaatgtactc aaatctcatc 2100  
 agcagcgtct acatcgtaaa aaacaattag agaatgaaat gatgcggggt ggattatctc 2160  
 aagatgccca ggatcaaatg agaaagatgc ttgcccacaa agaattctaat tacatccgtc 2220  
 ttaaaagggc taaaatggac aagtctatgt ttgtgaagat aaagacacta ggaataggag 2280  
 catttggtga agtctgtcta gcaagaaaag tagatactaa ggctttgtat gcaacaaaaa 2340  
 ctcttcgaaa gaaagatgtt ctcttcgaa atcaagtcgc tcatgttaag gctgagagag 2400  
 atatcctggc tgaagctgac aatgaatggg tagttcgtct atattattca ttccaagata 2460  
 aggacaattt atactttgta atggactaca ttctggggg tgatatgatg agcctattaa 2520  
 ttagaatggg catctttcca gaaagtctgg cacgattcta catagcagaa cttacctgtg 2580  
 cagttgaaag tgttcataaa atgggtttta ttcatagaga tattaacact gataatattt 2640  
 tgattgatcg tgatggatcat attaaattga ctgactttgg cctctgcact ggcttcagat 2700  
 ggacacacga ttctaagtac tatcagagtg gtgaccatcc acggcaagat agcatggatt 2760  
 tcagtaatga atggggggat ccctcaagct gtcgatgtgg agacagactg aagccattag 2820  
 agcggagagc tgcacgccag caccagcgt gtctagcaca ttctttgggt gggactccca 2880  
 attatattgc acctgaagtg ttgctacgaa caggatacac acagttgtgt gattgggtga 2940  
 gtgttggtgt tattcttttt gaaatgttg tgggacaacc tcctttcttg gcacaaacac 3000  
 cattagaaac acaaatgaag gtcacctgct gctatatata tcattggctc gagaagaaac 3060  
 tactgaacac cctgcgagag agaagcctag aaaagaaaga aagggccaaa aggttttgaa 3120

ctcttcatcc ctaatttgct acactgatca aaaccaagta agggctcctg aagtccatga 3180  
 gtctatcatc aatcagcaca aatgctatac tagtttgtaa ctgcggggtc agttgtgaag 3240  
 gggaaggaca gcagtcttat ccatattcca ggaagccaca gttaaactgct cga 3293

<210> 51  
 <211> 424  
 <212> DNA  
 <213> Homo sapiens

<400> 51  
 cctactctat tcagatattc tcagattcc taaagattag agatcatttc tcattctcct 60  
 aggagtactc acttcaggaa gcaaccagat aaaagagagg tgcaacggaa gccagaacat 120  
 tcctcctgga aattcaacct gtttcgcagt ttctcgagga atcagcattc agtcaatccg 180  
 ggccgggagc agtcatctgt ggtgaggctg attggctggg caggaacagc gccggggcgt 240  
 gggctgagca cagcgcttcg ctctctttgc cacaggaagc ctgagctcat tcgagtagcg 300  
 gctcttccaa gctcaaagaa gcagaggccg ctgttcgttt cctttaggtc tttccactaa 360  
 agtcggagta tcttcttcca agatttcacg tcttggtggc cgttccaagg agcgcgaggt 420  
 cggg 424

<210> 52  
 <211> 706  
 <212> DNA  
 <213> Homo sapiens

<400> 52  
 tgaactctga ctgtatgaga tgtaaatac tttttaatat ttgttttagat atgacattta 60  
 ttcaaagtta aaagcaaaca cttacagaat tatgaagagg tatctgttta acatttcctc 120  
 agtcaagttc agagtcttca gagacttcgt aattaaagga acagagtgag agacatcatc 180  
 aagtggagag aaatcatagt ttaaactgca ttataaattt tataacagaa ttaaagtaga 240  
 ttttaaaaga taaaatgtgt aattttgttt atattttccc atttggaactg taactgaactg 300  
 ccttgctaaa agattataga agtagcaaaa agtattgaaa tgtttgcata aagtgtctat 360  
 aataaaacta aactttcatg tgactggagt catcttgtcc aaactgcctg tgaatatatc 420  
 ttctctcaat tggaatattg tagataactt ctgctttaa aaagttttct ttaaataatc 480  
 ctactcattt ttgtgggaat ggttaagcag tttaaataat tcctgtgtat atgtctatca 540  
 catagggggtc taacagaaca atctggattc attatttcta ggacttgatc ctgctgatgc 600

tgaatttgca cattaagggtg tgtaacaac caaaacacag atcgatataa gaagtaagga 660  
 ggtggggaga ggcaaattat gatgtgctat gagttagatg tatagt 706

<210> 53  
 <211> 239  
 <212> DNA  
 <213> Homo sapiens

<400> 53  
 agtccgcggc gttccccggc tgcagccggg agggggccga ggagtgactg agccccgggc 60  
 tgtgcagtcc gacgccgact gaggcacgag cgggtgacgc tgggcctgca gcgcggagca 120  
 gaaagcagaa cccgcagagt cctccctgct gctgtgtgga cgacacgtgg gcacaggcag 180  
 aagtgggccc tgtgaccagc tgcactgggt tctgtgaagg aagctccagg actggcggg 239

<210> 54  
 <211> 641  
 <212> DNA  
 <213> Homo sapiens

<400> 54  
 tgaggcagct gctatcccca tctccctgcc tggcccccaa cctcagggtc cccaggggtc 60  
 tccctggctc cctcctccag gcctgcctcc cacttcaactg cgaagaccct cttgccacc 120  
 ctgactgaaa gtagggggct ttctggggcc tagcgatctc tcttggccta tccgctgcca 180  
 gccttgagcc ctggctgttc tgtggttctc ctgtcaccg cccatcaggg ttctcttctc 240  
 aactcagaga aaaatgctcc ccacagcgtc cctggcgagc gtgggctgga cttctacctg 300  
 cctcaagggt tgtgtatatt gtataggggc aactgtatga aaaattgggg aggagggggc 360  
 cgggcgcggt gctcacgct gtaatccag cactttggga ggccgaggcg ggtggatcac 420  
 gaggtcagga gatcgagacc atcctggcta acatggtgaa acccgtctc tactaaaaat 480  
 aaaaaaaaaa tttagccggg cgcggtggcg ggcacctgta gtcccagcta cttgggaggc 540  
 tgaggcagga gaatggtgtg aaccgggag cggaggttgc agtgagctga gatcgtgcta 600  
 ctgcactcca gcctggggga cagaaagaga ctccgtctca a 641

<210> 55  
 <211> 493  
 <212> DNA  
 <213> Homo sapiens

<400> 55  
 tttctgtgaa gcagaagtct gggaatcgat ctggaaatcc tctaatttt tactccctct 60

cccccgact cctgattcat tgggaagttt caaatcagct ataactggag agagctgaag 120  
 attgatggga tcgttgccctt atgcctttgt tttggtttta caaaaaggaa acttgacaga 180  
 ggatcatgct atacttaaaa aatacaacat cgcagaggaa gtagactcat attaaaaata 240  
 cttactaata ataacgtgcc tcatgaagta aagatccgaa aggaattgga ataaaaacttt 300  
 cctgcatctc aagccaaggg ggaaacacca gaatcaagtg ttccgcgtga ttgaagacac 360  
 cccctcgtcc aagaatgcaa agcacatcca ataaaagagc tggattataa ctctcttct 420  
 ttctctgggg gccgtggggg gggagctggg gcgagaggtg ccgttggtccc ccgttgcttt 480  
 tcctctggga ggg 493

<210> 56  
 <211> 5282  
 <212> DNA  
 <213> Homo sapiens

<400> 56  
 tgaagtcaac atgcctgccc caaacaata tgcaaaaggt tcactaaagc agtagaaata 60  
 atatgcattg tcagtgatgt tccatgaaac aaagctgcag gctgtttaag aaaaaataac 120  
 acacatataa acatcacaca cacagacaga cacacacaca cacaacaatt aacagtcttc 180  
 aggcaaaaag tcgaatcagc tatttactgc caaagggaaa tatcatttat tttttacatt 240  
 attaagaaaa aaagatttat ttatttaaga cagtcccatc aaaactcctg tctttggaaa 300  
 tccgaccact aattgccaaag caccgcttcg tgtggctcca cctggatgtt ctgtgcctgt 360  
 aaacatagat tcgctttcca tgttggtggc cggatcacca tctgaagagc agacggatgg 420  
 aaaaaggacc tgatcattgg ggaagctggc tttctggctg ctggaggctg gggagaaggt 480  
 gttcattcac ttgcatttct ttgccttggg ggctgtgata ttaacagagg gagggttcct 540  
 gtggggggaa gtccatgcct ccctggcctg aagaagagac tctttgcata tgactcacat 600  
 gatgcatacc tgggtgggagg aaaagagttg ggaacttcag atggacctag taccactga 660  
 gatttccacg ccgaaggaca gcgatgggaa aaatgccctt aaatcatagg aaagtatttt 720  
 ttttaagctac caattgtgcc gagaaaagca ttttagcaat ttatacaata tcatccagta 780  
 ccttaagccc tgattgtgta tattcatata ttttggtatc gcacccccca actcccaata 840  
 ctggctctgt ctgagtaaga aacagaatcc tctggaactt gaggaagtga acatttcggt 900  
 gacttccgca tcaggaaggc tagagttacc cagagcatca ggccgccaca agtgccctgct 960  
 tttaggagac cgaagtccgc agaacctgcc tgtgtcccag cttggagggc tggtcctgga 1020

actgagccgg ggcctcact ggcctcctcc agggatgac aacagggcag tgtggtctcc 1080  
 gaatgtctgg aagctgatgg agctcagaat tccactgtca agaaagagca gtagaggggt 1140  
 gtggctgggc ctgtcaccct ggggcccctcc aggtaggccc gttttcacgt ggagcatggg 1200  
 agccacgacc cttcttaaga catgtatcac tgtagaggga aggaacagag gccctgggcc 1260  
 cttcctatca gaaggacatg gtgaaggctg ggaacgtgag gagaggcaat ggccacggcc 1320  
 cattttggct gtagcacatg gcacgttggc tgtgtggcct tggcccacct gtgagtttaa 1380  
 agcaaggctt taaatgactt tggagagggt cacaatcct aaaagaagca ttgaagtgag 1440  
 gtgtcatgga ttaattgacc cctgtctatg gaattacatg taaaacatta tcttgtcact 1500  
 gtagtttgggt tttatttgaa aacctgacaa aaaaaaagt ccaggtgtgg aatatggggg 1560  
 ttatctgtac atcctggggc attaaaaaaa aaatcaatgg tggggaacta taaagaagta 1620  
 acaaaagaag tgacatcttc agcaaataaa ctaggaaatt ttttttctt ccagtttaga 1680  
 atcagccttg aaacattgat ggaataactc tgtggcatta ttgcattata taccatttat 1740  
 ctgtattaac tttggaatgt actctgttca atgtttaatg ctgtggttga tatttcgaaa 1800  
 gctgctttaa aaaaatacat gcatctcagc gtttttttgt ttttaattgt atttagttat 1860  
 ggcctataca ctatttgtga gcaaaggatg tcttttctg tttgagattt ttatctcttg 1920  
 attcttcaaa agcattctga gaaggatgaga taagccctga gtctcagcta cctaagaaaa 1980  
 acctggatgt cactggccac tgaggagctt tgtttcaacc aagtcattgt catttccacg 2040  
 tcaacagaat tgtttattgt gacagttata tctgttgtcc ctttgacctt gtttcttgaa 2100  
 ggtttcctcg tccctgggca attcgcatt taattcatgg tattcaggat tacatgcatg 2160  
 tttgggttaa cccatgagat tcattcagtt aaaaatccag atggcaaatg accagcagat 2220  
 tcaaactctat ggtgggttga cctttagaga gttgctttac gtggcctgtt tcaacacaga 2280  
 cccaccaga gccctcctgc cctccttccg cgggggcttt ctcatggctg tcttcaggg 2340  
 tottctgaa atgcagtggg gcttacgctc caccaagaaa gcaggaaacc tgtggtatga 2400  
 agccagacct ccccgcggg cctcaggga cagaatgac agaccttga atgattctaa 2460  
 tttttaagca aaatattatt ttatgaaagg tttacattgt caaagtgtg aatatggaat 2520  
 atccaatcct gtgctgctat cctgccaaaa tcattttaat ggagtcagtt tgcagtatgc 2580  
 tccaogtggg aagatcctcc aagctgcttt agaagtaaca atgaagaacg tggacgcttt 2640  
 taatataaag cctgttttgt cttctgttgt tgttcaaacg ggattcacag agtatttgaa 2700

aaatgtatat atattaagag gtcacggggg ctaattgctg gctggctgcc ttttgctgtg 2760  
 gggttttgtt acctggtttt aataacagta aatgtgccca gctccttggc cccagaactg 2820  
 tacagtattg tggtgcact tgccttaaga gtagttgatg ttgcattttc cttattgtta 2880  
 aaaacatgtt agaagcaatg aatgtatata aaagcctcaa ctagtcattt ttttctctc 2940  
 ttcttttttt toattatata taattatttt gcagttgggc aacagagaac catccctatt 3000  
 ttgtattgaa gagggattca catctgcac ttaactgctc tttatgaatg aaaaaacagt 3060  
 cctctgtatg tactcctctt tacactgggc agggtcagag ttaaataagag tatatgcact 3120  
 ttccaaattg gggacaaggg ctctaaaaaa agcccaaaaa ggagaagaac atctgagaac 3180  
 ctctcgggc ctcccagtc ctcgctgcac aaatactccg caagagaggc cagaatgaca 3240  
 gctgacaggg tctatggcca tcgggtcgtc tccgaagatt tggcaggggc agaaaactct 3300  
 ggcaggctta agatttgaa taaagtcaca gaatcaagga agcacctcaa tttagttcaa 3360  
 acaagacgcc aacattctct ccacagctca cttacctctc tgtgttcaga tgtggccttc 3420  
 catttatatg tgatctttgt tttattagta aatgcttata atctaaagat gtagctctgg 3480  
 cccagtggga aaaattagga agtgattata aatcgagagg agttataata atcaagatta 3540  
 aatgtaaata atcagggcaa tccaacaca tgtctagctt tcacctccag gatctattga 3600  
 gtgaacagaa ttgcaaatag tctctatttg taattgaact tatcctaaaa caaatagttt 3660  
 ataaatgtga acttaaaact taattaattc caactgtact ttttaaggcag tggctgtttt 3720  
 tagactttct tatcacttat agttagtaat gtacacctac tctatcagag aaaaacagga 3780  
 aaggctcgaa atacaagcca ttctaaggaa attagggagt cagttgaaat tctattctga 3840  
 tcttattctg tgggtgtctt tgcagcccag acaaatgtgg ttacacactt ttaagaaat 3900  
 acaattctac attgtcaagc ttatgaaggc tccaatcaga tttttattgt tattcaattt 3960  
 ggatctttca gggatttttt ttttaaatta ttatgggaca aaggacattt gttggagggg 4020  
 tgggaggagg gaacaatttt taaatataaa acattcccaa gtttgatca gggagttgga 4080  
 agttttcaga ataaccagaa ctaagggtat gaaggacctg tattggggtc gatgtgatgc 4140  
 ctctgcgaag aaccttgtgt gacaaatgag aaacattttg aagtttgtgg tacgacctt 4200  
 agattccaga gacatcagca tggctcaaag tgcagctccg tttggcagtg caatggtata 4260  
 aatttcaagc tggatatgtc taatgggtat ttaacaata aatgtgcagt ttaactaac 4320  
 aggatattta atgacaacct tctggttggg agggacatct gtttctaaat gtttattatg 4380  
 tacaatacag aaaaaattt tataaaatta agcaatgtga aactgaattg gagagtgata 4440

atacaagtcc ttagtcttta ccagtgat cattctgttc catgtctttg gacaaccatg 4500  
 accttggaac atcatgaaat atgcatctca ctggatgcaa agaaaatcag atggagcatg 4560  
 aatgggtactg tacoggttca tctggactgc ccagaaaaa taacttcaag caaacatcct 4620  
 atcaacaaca aggttgttct gcataccaag ctgagcacag aagatgggaa cactgggtgga 4680  
 ggatggaaaag gctcgctcaa tcaagaaaat tctgagacta ttaataaata agactgtagt 4740  
 gtagatactg agtaaatcca tgcacctaaa ccttttgga aatctgccgt gggccctcca 4800  
 gatagctcat ttcattaagt ttttccctcc aaggtagaat ttgcaagagt gacagtggat 4860  
 tgcatttctt ttgggggaagc tttcttttgg tggttttgtt tattatacct tcttaagttt 4920  
 tcaaccaagg tttgcttttg ttttgagtta ctgggggttat ttttgtttta aataaaaata 4980  
 agtgtacaat aagtgttttt gtattgaaag cttttgttat caagattttc atacttttac 5040  
 cttccatggc tctttttaag attgatactt ttaagaggtg gctgatattc tgcaacactg 5100  
 tacacataaa aaatacggta aggatacttt acatgggttaa ggtaaagtaa gtctccagtt 5160  
 ggccaccatt agctataatg gcactttgtt tgtgttgttg gaaaaagtca cattgccatt 5220  
 aaactttcct tgtctgtcta gttaatattg tgaagaaaaa taaagtacag tgtgagatac 5280  
 tg 5282

<210> 57  
 <211> 117  
 <212> DNA  
 <213> Homo sapiens

<400> 57  
 attcggggcg agggaggagg aagaagcgga ggaggcggt cccgctcgca gggccgtgca 60  
 cctgcccgcg cgcccgcctg ctgctcgcc cgccgcgcg cgctgcgcg cgccagc 117

<210> 58  
 <211> 430  
 <212> DNA  
 <213> Homo sapiens

<400> 58  
 tgatccaggg agccccacc atccgggggg accccgagtg tcattctctt tacaatgagc 60  
 agcaggaggg ttgcgggggtg cacaccagc ggatgcagta gaccgcagcc agccggtgcc 120  
 tggcgccct gcccccgcg cctctccaaa caccggcaga aaacggagag tgcttgggtg 180  
 gtgggtgctg gaggattttc cagttctgac acacgtattt atatttgga agagaccagc 240

accgagctcg gcacctcccc ggcctctctc ttcccagctg cagatgccac acctgctcct 300  
 tcttgctttc cccgggggag gaaggggggt gtggtcgggg agctggggta cagggtttggg 360  
 gagggggaag agaaatTTTT atttttgaac cctgtgtcc cttttgcata agattaaagg 420  
 aaggaaaagt 430

<210> 59  
 <211> 192  
 <212> DNA  
 <213> Homo sapiens

<400> 59  
 tcctaggcgg cggccgcggc ggcggaggca gcagcggcgg cggcagtggc ggcggcgaag 60  
 gtggcggcgg ctcgccagt actcccgcc cccgccattt cggactggga gcgagcgcgg 120  
 cgcaggcact gaaggcggcg gcggggccag aggctcagcg gctcccagggt gcgggagaga 180  
 ggcttgcctga aa 192

<210> 60  
 <211> 4172  
 <212> DNA  
 <213> Homo sapiens

<400> 60  
 taaatacaat ttgtactttt ttcttaaggc atactagtac aagtggtaat ttttgtacat 60  
 tacactaaat tattagcatt tgtttttagca ttacctaat tttttcctgc tccatgcaga 120  
 ctgttagctt ttaccttaa tgcttatttt aaaatgacag tggaagtttt tttttcctcg 180  
 aagtgcagat attcccagag ttttggtttt tgaactagca atgcctgtga aaaagaaact 240  
 gaatacctaa gatttctgtc ttgggggttt tggtgcatgc agttgattac ttottatttt 300  
 tottaocaaag tgtgaatggt ggtgtgaaac aaattaatga agcttttgaa tcatccctat 360  
 totgtgtttt atctagtcac ataaatggat taattactaa tttcagttga gaccttctaa 420  
 ttggtttttt ctgaaacatt gagggacaca aatttatggg cttcctgatg atgattcttc 480  
 taggcacat gtcctatagt ttgtcatccc tgatgaatgt aaagttacac tgttcacaaa 540  
 ggttttgtct cttttccact gctattagtc atggctactc tccccaaaat attatatattt 600  
 ttctataaaa agaaaaaaat ggaaaaaaat tacaaggcaa tggaaaactat tataaggcca 660  
 tttccttttc acattagata aattactata aagaactcta atagcttttt cctgttaagg 720  
 cagacccagat atgaatggga ttattatagc aaccattttg gggctatatt tacatgctac 780



taaattttta taataattga aaagatttta acaagtataa aaaaattctc ataggaatta	840
aatgtagtct cctgtgtca gactgctctt tcatagtata actttaaatc ttttcttcaa	900
cttgagtctt tgaagatagt tttaattctg cttgtgacat taaaagatta tttgggccag	960
ttatagctta ttaggtgttg aagagaccaa ggttgcaagc caggccctgt gtgaaccttg	1020
agctttcata gagagtttca cagcatggac tgtgtgcccc acggtcatcc gagtgggtgt	1080
acgatgcatt ggttagtcaa aaatggggag ggactagggc agtttgata gctcaacaag	1140
atacaatctc actctgtggt ggtcctgctg acaaatcaag agcattgctt ttgtttctta	1200
agaaaacaaa ctctttttta aaaattactt ttaaatatta actcaaaagt tgagattttg	1260
gggtgggtgt gtgccaagac attaatTTTT tttttaaaca atgaagtga aaagttttac	1320
aatctctagg tttggctagt tctcttaaca ctggttaaat taacattgca taaacaottt	1380
tcaagtctga tccatattta ataatgcttt aaaataaaaa taaaaacaat ccttttgata	1440
aatttaaaat gttacttatt ttaaaataaa tgaagtgaga tggcatgggt aggtgaaagt	1500
atcactggac taggttgttg gtgaacttag ttctagatag gtgtctttta ggactctgat	1560
tttgaggaca tcaattacta tccatttctt catgttaaaa gaagtcatct caaactctta	1620
gttttttttt tttaactat gtgatttata ttccatttac ataaggatac acttatttgt	1680
caagctcagc acaatctgta aatttttaac ctatgttaca ccatcttcag tgccagtctt	1740
gggcaaaatt gtgcaagagg tgaagtttat atttgaatat ccattctcgt tttaggactc	1800
ttcttccata ttagtgtcat cttgcctccc taccttccac atgccccatg acttgatgca	1860
gttttaatac ttgtaattcc cctaaccata agatttactg ctgctgtgga tatctocatg	1920
aagttttccc actgagtcac atcagaaatg ccctacatct tattttctc agggctcaag	1980
agaatctgac agataccata aagggatttg acctaatcac taattttcag gtggtggtg	2040
atgctttgaa catctctttg ctgccaatc cattagcgac agtaggattt ttcaacctg	2100
gtatgaatag acagaacct atccagtgga aggagaattt aataaagata gtgcagaaag	2160
aattccttag gtaatctata actaggacta ctcttggtta cagtaatata ttccattggt	2220
ttagtaacca gaaatcttca tgcaatgaaa aatactttaa ttcatgaagc ttactttttt	2280
ttttttgtg tcaagatctc gctcttgtca ccagggtgg aatgcagtgg cgccatctca	2340
gctcactgca accttccatc ttcccagggt caagcgatc togtgcctcg gcctcctgag	2400
tagctgggat tacaggcgtg tgcactacac tcaactaatt tttgtatttt taggagagac	2460
gggtttcac ctgttggcca ggtggtctc gaactcctga cctcaagtga ttcaccacc	2520

ttggcctcat aaacctgttt tgcagaactc atttattcag caaatattta ttgagtgcc 2580  
 accagatgcc agtcaccgca caaggcactg ggtatatggt atccccaaac aagagacata 2640  
 atcccggtcc ttaggtactg ctagtgtggt ctgtaatatc ttactaaggc ctttggtata 2700  
 cgaccagag ataacacgat gcgtatttta gttttgcaaa gaaggggttt ggtctctgtg 2760  
 ccagctctat aattgttttg ctacgattcc actgaaactc ttogatcaag ctactttatg 2820  
 taaatcactt cattgtttta aaggaataaa cttgattata ttgttttttt atttggcata 2880  
 actgtgatcc ttttaggaca attactgtac acattaaggt gtatgtcaga tattcatatt 2940  
 gacccaaatg tgtaatatcc cagttttctc tgcataagta attaaaatat acttaaaaat 3000  
 taatagtttt atctgggtac aaataaacag tgccgaact agttcacaga caagggaac 3060  
 ttctatgtaa aaatcactat gatttctgaa ttgctatgtg aaactacaga tctttggaac 3120  
 actgtttagg taggggtgta agacttgaca cagtacctg tttctacaca gagaaagaaa 3180  
 tggccatact tcaggaactg cagtgcctat gaggggatat ttaggcctct tgaatttttg 3240  
 atgtagatgg gcattttttt aaggtagtgg ttaattacct ttatgtgaac tttgaatgg 3300  
 ttaacaaaag atttgttttt gtagagattt taaaggggga gaattctaga aataaatgtt 3360  
 acctaattat tacagcctta aagacaaaaa tccttggtga agttttttta aaaaagact 3420  
 aaattacata gacttaggca ttaacatgtt tgtggaagaa tatagcagac gtatattgta 3480  
 tcatttgagt gaatgttccc aagtaggcat tctaggtctt atttaactga gtcacactgc 3540  
 ataggaattt agaacctaac ttttataggt tatcaaaact gttgtcacca ttgcacaatt 3600  
 ttgtccta atatacatag aaactttgtg gggcatgtta agttacagtt tgcacaagtt 3660  
 catctcattt gtattccatt gatttttttt tttcttctaa acattttttt ttcaaaacag 3720  
 tatatataac tttttttagg ggattttttt tagacagcaa aaaactatct gaagatttcc 3780  
 atttgtcaaa aagtaatgat ttcttgataa ttgtgtagtg aatgtttttt agaaccagc 3840  
 agttacctg aaagctgaat ttatatntag taacttctgt gtttaactg gatagcatga 3900  
 attctgcatt gagaaactga atagctgtca taaaatgctt tctttcctaa agaaagatac 3960  
 tcacatgagt tcttgaagaa tagtcataac tagattaaga tctgtgtttt agtttaatag 4020  
 tttgaagtgc ctgtttggga taatgatag taatttagat gaatttaggg gaaaaaaaag 4080  
 ttatctgcag ttatgttgag ggccatctc tccccacaca cccccacaga gctaactggg 4140  
 ttacagtgtt ttatccgaaa gtttccaatt cc 4172

<210> 61  
 <211> 238  
 <212> DNA  
 <213> Homo sapiens

<400> 61  
 ccattgtgct ggaaaggcgc gcaacggcgg cgacggcggc gacccaccg cgcacccctgc 60  
 caggcctccg cgcccagccg cccacggcgc cccgcgcccc ggcggccgac cctttcttcg 120  
 cgcccccgcc cctcgccccg ccaggccccc ttgcggcca cccgccaggc cccgcgcccg 180  
 cccgccccgc gccagggacc ggcccgccgc ccgcaggccg cccgccgcc gcgcgcc 238

<210> 62  
 <211> 547  
 <212> DNA  
 <213> Homo sapiens

<400> 62  
 ggccccgcag ctctggccac agggacctct gcagtgcgcc ctaagtgacc cggacaattc 60  
 cgagggggcc atcaccgcct gtgtatataa cgtttcgggt attactctgc tacacgtagc 120  
 ctttttactt ttgggggttt gtttttgttc tgaactttcc tgttaccttt tcagggtga 180  
 tgtcacatgt aggtggcgtg tatgagtgga gacgggcctg ggtcttgggg actggagggc 240  
 aggggtcctt ctgcccttg ggtcccaggg tgctctgcct gctcagccag gcctctcctg 300  
 ggagccactc gccagagac tcagcttggc caacttgggg ggctgtgtcc acccagcccg 360  
 cccgtcctgt gggctgcaca gctcaccttg ttcctcctg ccccggttcg agagccgagt 420  
 ctgtgggcac tctctgcctt catgcacctg tcctttctaa cagtcgcct tcaactgtaa 480  
 tcacaacatc ctgactccgt catttaataa agaaggaaca tcaggcatgc taaaaaaaaa 540  
 aaaaaaa 547

<210> 63  
 <211> 102  
 <212> DNA  
 <213> Homo sapiens

<400> 63  
 gaattccggc aaacatgagg cagctgccag ccggcctggg cagtcttgtc tgcctcggct 60  
 gtgaagtggg gaggtggca acagttttct tcagcgcca gg 102

<210> 64  
 <211> 2017

&lt;212&gt; DNA

&lt;213&gt; Homo sapiens

&lt;400&gt; 64

gacacgtcca aaggagtgca tggccacagc cacctccacc cccaagaaac ctccatcctg	60
ccaggagcag cctccaagaa actttttaaaa aatagatttg caaaaagtga acagattgct	120
acacacacac acacacacac acacacacac acacacagcc attcatcttg gctggcagag	180
gggacagagt tcagggaggg gctgagttct gctagggggc ggtccagag gccccagcca	240
gcccttccca ggccagcgag gcgaggctgc ctctgggtga gtggctgaca gagcaggtct	300
gcaggccacc agctgctgga tgtaaccaag aaggggctcg agtgccctgc aggaggggtcc	360
aatcctccgg tccacctcg tcccgttcat ccattctgct ttcttgccac acagtggccg	420
gccaggtctc ccctgggtct ctccccgtag ccactctctg ccactacct atgcttctag	480
aaagcccctc acctcaggac ccagaggac cagctggggg gcagggggga gagggggtaa	540
tggaggccaa gcctgcagct ttctggaaat tcttccctgg ggggtccagt atccccgtct	600
actccactga cctggaagag ctgggtacca ggccaccac tgtggggcaa gcctgagtgg	660
tgaggggcca ctggcatcat tctccctcca tggcaggaag gcgggggatt tcaagtttag	720
ggattgggtc gtggtggaga atctgagggc actctgccag ctccacaggt ggatgagcct	780
ctccttgccc cagtctggt tcagtgggaa tgcagtgggt ggggctgtac acacctcca	840
gcacagactg ttccctccaa ggtcctctta ggtcccgggg aggaacgtgg ttcagagact	900
ggcagccagg gagcccgggg cagagctcag aggagtctgg gaaggggctg gtccctctctc	960
ttcctgtagt gcccctccca tggcccagca gcttggtga gcccctctcc tgaagcagct	1020
gtgcgcgctc cctctgcctt gcacaaaaag cacaagacat tccttagcag ctacgcgag	1080
ccctagtggg agcccagcac actgcttctc ggaggccagg ccctcctgct ggctgagctt	1140
gggcccgtg gccccaatat ggtggccctg ggaagaggg cttgggggtc tgcctctgtc	1200
ctgggatcag tggggcccca aagcccagcc cggtgacca acattcaaaa gcacaaacc	1260
tggggactct gcttggtgt cccctccatc tggggatgga gaatgcagcc caaagctgga	1320
gccaatggtg agggctgaga gggctgtggc tgggtggtca gcagaaacc caggaggaga	1380
gagatgctgc tccgcctga ttggggcctc accagaagg aaccgggtcc cagccgcatg	1440
gcccctccag gaacattccc acataatata ttccatcaca gccagcccag ctccactcag	1500
ggctggcccg gggagtcccc gtgtgcccc agaggctagc ccaggggtga gcagggccct	1560
cagaggaaag gcagtatggc ggaggccatg ggggcccctc ggcattcaca cacagcctgg	1620

cctccccctgc ggagctgcat ggacgcctgg ctccaggctc caggctgact ggggcctctg 1680  
 cctccaggag ggcacagct ttccctggct cagggatctt ctccctcccc tcaccgctg 1740  
 cccagccctc ccagctgatg tcactctgcc tctaagccaa ggctcagga gagcatcacc 1800  
 accacaccct gcggccttgc cttggggcca gactggctgc acagcccaac caggaggggt 1860  
 ctgcctccca cgctgggaca cagaccggcc gcatgtctgc atggcagaag cgtctccctt 1920  
 gccacggcct gggaggggtg ttctgtttct cagcatccac taatattcag tcctgtatat 1980  
 ttttaataaaa taaacttgac aaaggaaaaa aaaaccg 2017

<210> 65  
 <211> 97  
 <212> DNA  
 <213> Homo sapiens

<400> 65  
 gtccaggaac tctcagcag cgcctccttc agctccacag ccagacgcc tcagacagca 60  
 aagcctaccc ccgcgccgcg cctgcccgc cgctgcg 97

<210> 66  
 <211> 1474  
 <212> DNA  
 <213> Homo sapiens

<400> 66  
 aagtctaattg atcatattta tttattttata tgaaccatgt ctattaattt aattatttta 60  
 taatattttat attaaactcc ttatgttact taacatcttc tgtaacagaa gtcagtactc 120  
 ctgttgcgga gaaaggagtc atacttgtga agacttttat gtcactactc taaagatttt 180  
 gctgttgcgtg ttaagtttgg aaaacagttt ttattctgtt ttataaacca gagagaaatg 240  
 agttttgacg tcttttttact tgaatttcaa cttatattat aaggacgaaa gtaaagatgt 300  
 ttgaataactt aaacactatc acaagatgcc aaaatgctga aagtttttac actgtcgatg 360  
 tttccaatgc atcttccatg atgcattaga agtaactaat gtttgaaatt ttaaagtact 420  
 tttgggtatt tttctgtcat caaacaaaac aggtatcagt gcattattaa atgaatattt 480  
 aaattagaca ttaccagtaa tttcatgtct acttttttaa atcagcaatg aaacaataat 540  
 ttgaaatttc taaattcata gggtagaatc acctgtaaaa gcttgtttga tttcttaaag 600  
 ttattaaact tgtacatata ccaaaaagaa gctgtcttgg atttaaactt gtaaaatcag 660  
 atgaaatttt actacaattg cttgttaaaa tattttataa gtgatgttcc tttttcacca 720

```

agagtataaa ccttttttagt gtgactgtta aaacttcott tttaatcaaa atgocaaatt 780
tattaagggtg gtggagccac tgcagtgtta tctcaaaata agaatatcct gttgagatat 840
tccagaatct gtttatatgg ctggtaacat gtaaaaaccc cataaccccg ccaaaagggg 900
tcctaccott gaacataaag caataaccaa aggagaaaag cccaaattat tggttccaaa 960
tttagggttt aaactttttg aagcaaactt ttttttagcc ttgtgactg cagacctggg 1020
actcagattt tgctatgagg ttaatgaagt accaagctgt gcttgaataa cgatatgttt 1080
tctcagattt tctgttgtac agtttaattt agcagtcct atcacattgc aaaagtagca 1140
atgacctcat aaaatacctc ttcaaaatgc ttaaatcoat ttcacacatt aattttatct 1200
cagtcttgaa gccaatcag taggtgcatt ggaatcaagc ctggctacct gcctgctgtt 1260
ccttttcttt tcttctttta gccattttgc taagagacac agtcttctca aacacttcgt 1320
ttctcctatt ttgttttact agttttaaga tcagagttca ctttcttttg actctgccta 1380
tattttctta cctgaacttt tgcaagtttt caggtaaacc tcagctcagg actgctattt 1440
agctcctctt aagaagatta aaaaaaaaaa aaaa 1474

```

<210> 67  
 <211> 99  
 <212> DNA  
 <213> Homo sapiens

```

<400> 67
gcgccccgcc cccacccctc gcagcacccc gcgccccgog ccctcccagc cgggtccagc 60
cggagccatg gggccggagc cgcagtgagc accatggag 99

```

<210> 68  
 <211> 614  
 <212> DNA  
 <213> Homo sapiens

```

<400> 68
tgaaccagaa ggccaagtcc gcagaagccc tgatgtgtcc tcaggagca gggaaggcct 60
gacttctgct ggcatcaaga ggtgggaggg cctccgacc acttcaggg gaacctgcca 120
tgccaggaa cgttcctaag gaaccttctt tctgtctga gttcccagat ggctggaagg 180
ggtccagcct cgttggaaga ggaacagcac tggggagtct ttgtggattc tgaggccctg 240
cccaatgaga ctctagggtc cagtggatgc cacagcccag cttggccctt tccttcaga 300
tcctgggtac tgaaagcctt agggaagctg gcctgagagg ggaagcggcc ctaaggaggt 360
gtctaagaac aaaagcgacc catcagaga ctgtccctga aacctagtag tgcccccat 420

```

gaggaaggaa cagcaatggt gtcagtatcc aggctttgta cagagtgctt ttctgttttag 480  
 tttttacttt ttttgttttg tttttttaaa gacgaaataa agacccaggg gagaatgggt 540  
 gttgtatggg gaggaagtg tgggggggtcc ttctccacac ccactttgtc catttgcaaa 600  
 tatatttttg aaaa 614

<210> 69  
 <211> 36  
 <212> DNA  
 <213> Artificial

<220>  
 <223> Description of Artificial Sequence: Primer

<400> 69  
 aaagtcgacg taatcgcgga ggcttggggc agccgg 36

<210> 70  
 <211> 30  
 <212> DNA  
 <213> Artificial

<220>  
 <223> Description of Artificial Sequence: Primer

<400> 70  
 tttgcgactg gtcagctgcg ggatcccaag 30

<210> 71  
 <211> 33  
 <212> DNA  
 <213> Artificial

<220>  
 <223> Description of Artificial Sequence: Primer

<400> 71  
 aagtcgacgt aagagctcca gagagaagtc gag 33

<210> 72  
 <211> 33  
 <212> DNA  
 <213> Artificial

<220>  
 <223> Description of Artificial Sequence: Primer

<400> 72  
 aaacccgggc agcaaggcaa ggctccaatg cac 33

<210> 73  
 <211> 39  
 <212> DNA  
 <213> Artificial

<220>  
 <223> Description of Artificial Sequence: Primer

<400> 73  
 gccgggcagg aggaaggagc ctccctcagg gtttcggga 39

<210> 74  
 <211> 30  
 <212> DNA  
 <213> Artificial

<220>  
 <223> Description of Artificial Sequence: Primer

<400> 74  
 ctgcactaga gacaaagacg tgatgttaat 30

<210> 75  
 <211> 66  
 <212> DNA  
 <213> Artificial

<220>  
 <223> Description of Artificial Sequence: Polylinker

<400> 75  
 gaacaaatgt cgacgggggc ccctagcaga tctagcgctg gatcccccg ggagctcaug 60  
 gaagac 66

<210> 76  
 <211> 30  
 <212> DNA  
 <213> Artificial

<220>  
 <223> Description of Artificial Sequence: Primer

<400> 76  
 cgggtgtggg cgcgttattt atcggagttg 30

<210> 77  
 <211> 30  
 <212> DNA  
 <213> Artificial

<220>



<223> Description of Artificial Sequence: Primer

<400> 77

ttggcgaaga atgaaaatag ggttggtact

30

<210> 78

<211> 22

<212> DNA

<213> Artificial

<220>

<223> Description of Artificial Sequence: Primer

<400> 78

ggtgaaggtc ggagtcaacg ga

22

<210> 79

<211> 21

<212> DNA

<213> Artificial

<220>

<223> Description of Artificial Sequence: Primer

<400> 79

gagggatctc gctcctggaa g

21

<210> 80

<211> 55

<212> DNA

<213> Artificial

<220>

<223> Description of Artificial Sequence: Primer

<400> 80

aaagtcgacg taaccgccag atttgaatcg cgggacccgt tggcagaggt ggcg

55

<210> 81

<211> 54

<212> DNA

<213> Artificial

<220>

<223> Description of Artificial Sequence: Primer

<400> 81

aaaggatccg ggcaacgtcg gggcacccat gccgcgcgcg ccacctctgc caac

54

<210> 82

<211> 40

<212> DNA

<213> Artificial

<220>

<223> Description of Artificial Sequence: Primer

<400> 82  
aaagcggcgcg cggcctctgc cggagctgcc tgggtcccaga 40

<210> 83  
<211> 37  
<212> DNA  
<213> Artificial

<220>

<223> Description of Artificial Sequence: Primer

<400> 83  
aaatctagac tcaggaacag ccgagatgac ctccaga 37

<210> 84  
<211> 67  
<212> DNA  
<213> Artificial

<220>

<223> Description of Artificial Sequence: Primer

<400> 84  
ctagaagctt agggccgcgcg atccgcgcgc ggttcgcgcg gcgcggatcc gcggtagcaa 60  
gttagtc 67

<210> 85  
<211> 68  
<212> DNA  
<213> Artificial

<220>

<223> Description of Artificial Sequence: Primer

<400> 85  
gactaagctt gctaccgcgcg atccgcgcgc ggccaaccgc gcgcggatcc gcggccctaa 60  
gcttctag 68

<210> 86  
<211> 32  
<212> DNA  
<213> Artificial

<220>

<223> Description of Artificial Sequence: Primer

<400> 86  
 caagaagctt gcgcccggcc cccacccct cg 32

<210> 87  
 <211> 31  
 <212> DNA  
 <213> Artificial

<220>  
 <223> Description of Artificial Sequence: Primer

<400> 87  
 agcccattgt gctcactgag gctccggccc c 31

<210> 88  
 <211> 22  
 <212> DNA  
 <213> Artificial

<220>  
 <223> Description of Artificial Sequence: Primer

<400> 88  
 agactctgaa ccagaaggcc aa 22

<210> 89  
 <211> 36  
 <212> DNA  
 <213> Artificial

<220>  
 <223> Description of Artificial Sequence: Primer

<400> 89  
 ctccgtacca gttttccaaa atatatttgc aaatgg 36

<210> 90  
 <211> 58  
 <212> DNA  
 <213> Artificial

<220>  
 <223> Description of Artificial Sequence: Primer

<400> 90  
 cccaagcttc gcgcccggcc cccacccct cgcagcacc cgcgcccgc gccctccc 58

<210> 91  
 <211> 61  
 <212> DNA  
 <213> Artificial

&lt;220&gt;

&lt;223&gt; Description of Artificial Sequence: Primer

&lt;400&gt; 91

ggcccatggtg ctccggctgg acccggctgg gaccggctg ggagggcgcg ggagggcgcg 60

g

61

&lt;210&gt; 92

&lt;211&gt; 7008

&lt;212&gt; DNA

&lt;213&gt; Artificial

&lt;220&gt;

&lt;223&gt; Description of Artificial Sequence: Expression Vector

&lt;400&gt; 92

gacggatcgg gagatctccc gatccctat ggtgcactct cagtacaatc tgctctgatg 60

ccgcatagtt aagccagtat ctgctccctg cttgtgtgtt ggaggtcgct gagtagtgcg 120

cgagcaaaat ttaagctaca acaaggcaag gcttgaccga caattgcatg aagaatctgc 180

ttagggtttag gcgttttgcg ctgcttcgcg atgtacgggc cagatatagc cgttgacatt 240

gattattgac tagttattaa tagtaatcaa ttacggggtc attagttcat agcccatata 300

tggagttccg cgttacataa cttacggtaa atggcccgcc tggctgaccg cccaacgacc 360

cccgccatt gacgtcaata atgacgtatg ttcccatagt aacgccataa gggactttcc 420

attgacgtca atgggtggag tatttacggt aaactgccca cttggcagta catcaagtgt 480

atcatatgcc aagtacgccc cctattgacg tcaatgacgg taaatggccc gcctggcatt 540

atgccagta catgacctta tgggactttc ctacttggca gtacatctac gtattagtca 600

tcgctattac catggtgatg cggttttggc agtacatcaa tgggcgtgga tagcggtttg 660

actcacgggg atttccaagt ctccacccca ttgacgtcaa tgggagtttg ttttggcacc 720

aaaatcaacg ggactttcca aaatgtcgta acaactccgc cccattgacg caaatgggog 780

gtaggcgtgt acgggtggag gtctatataa gcagagctct ctggctaact aagctttcgg 840

cgcgccgagg taccatggga tccgaagacg ccaaaaacat aaagaaaggc ccggcgccat 900

tctatcctct agaggatgga accgctggag agcaactgca taaggctatg aagagatacg 960

ccctggttcc tggaacaatt gcttttacag atgcacatat cgaggtgaac atcacgtaac 1020

cggaataactt cgaaatgtcc gttcggtttg cagaagctat gaaacgatat gggtgaata 1080

caaatcacag aatcgtcgta tgcagtgaat actctcttca attctttatg ccggtgttgg 1140

gcgcgttatt tatcgaggtt gcagttgcgc ccgcgaacga cttttataat gaacgtgaat 1200

tgctcaacag	tatgaacatt	tgcgagccta	cgcgtagtgtt	tgtttccaaa	aaggggttgc	1260
aaaaaatttt	gaacgtgcaa	aaaaaattac	caataatcca	gaaaattatt	atcatggatt	1320
ctaaaacgga	ttaccagga	tttcagtoga	tgtacacgtt	cgtcacatct	catctacctc	1380
cgggttttaa	tgaatacgat	tttgtaccag	agtcotttga	tcgtgacaaa	acaattgcac	1440
tgataatgaa	ttcctctgga	tctactgggt	tacctaaggg	tgtggccctt	ccgcatagaa	1500
ctgcctgcgt	cagattctcg	catgccagag	atcctatttt	tggcaatcaa	atcattccgg	1560
atactgcgat	tttaagtgtt	gttccattcc	atcacggttt	tggaatgttt	actacactcg	1620
gatatttgat	atgtggattt	cgagtcgtct	taatgtatag	atttgaagaa	gagctgtttt	1680
tacgatccct	tcaggattac	aaaattcaaa	gtgcgttgct	agtaccaacc	ctattttcat	1740
tcttgcgcaa	aagcactctg	attgacaaat	acgatttata	taatttacac	gaaattgctt	1800
ctggggggcg	acctctttcg	aaagaagtcg	gggaagcggg	tgcaaaacgc	ttccatcttc	1860
cagggatagc	acaaggatat	gggctcactg	agactacata	agctattctg	attacaccgg	1920
agggggatga	taaaccgggc	gcggtcggta	aagttgttcc	attttttgaa	gcgaaggttg	1980
tggatctgga	taccgggaaa	acgctgggcg	ttaatcagag	aggcgaatta	tgtgtcagag	2040
gacctatgat	tatgtccggt	tatgtaaaca	atccggaagc	gaccaacgcc	ttgattgaca	2100
aggatggatg	gctacattct	ggagacatag	cttactggga	cgaagacgaa	cacttcttca	2160
tagttgaccg	cttgaagtct	ttaattaaat	acaaaggata	tcaggtggcc	cccgtgaat	2220
tggaatcgat	attgttacia	caccccaaca	tcttcgacgc	gggogtggca	ggtcttcccg	2280
acgatgacgc	cggtgaaact	cccgcgcgcg	ttgttgtttt	ggagcacgga	aagacgatga	2340
cggaaaaaga	gatcgtggat	tacgtcgcca	gtcaagtaac	aaccgcgaaa	aagttgcgcg	2400
gaggagtgtg	gtttgtggac	gaagtaccga	aaggtcttac	cggaaaactc	gacgcaagaa	2460
aaatcagaga	gatcctcata	aaggccaaga	agggcgga	gtccaaattg	cgcgccgct	2520
aactcgagaa	taaaatgagg	aaattgcata	gcattgtctg	agtaggtgtc	attctattct	2580
gggggggtgg	gtggggcagg	acagcaaggg	ggaggattgg	gaagacaata	gcaggcatgc	2640
tggggatgcg	gtgggctcta	tggcttctga	ggcggaaaga	accagctggg	gctctagggg	2700
gtatccccac	gcgccttgta	gcggcgcatt	aagcgcggcg	ggtgtggtgg	ttacgcgcag	2760
cgtgaccgct	acaattgcca	gcgccttagc	gcccgtcctt	ttegttttct	tcccttcctt	2820
tctcgccacg	ttcgccggct	ttccccgtca	agctctaaat	cgggggctcc	cttaggggtt	2880

ccgatttagt gctttacggc acctcgaccc caaaaaactt gattaggggtg atgggttcacg 2940  
 tagtgggcca tcgccctgat agacgggtttt tcgccctttg acgttggagt ccacgttctt 3000  
 taatagtgga ctcttggttc aaactggaac aacactcaac cctatctcgg tctattcttt 3060  
 tgattttataa gggatttttg cgatttcggc ctattgggtta aaaaatgagc tgatttaaca 3120  
 aaaatttaac gcgaattaat tctgtggaat gtgtgtcagt taggggtgtg aaagtcccca 3180  
 ggctccccag caggcagaag tatgcaaagc atgcatctca attagtcagc aaccagggtg 3240  
 ggaaagtccc caggctcccc agcaggcaga agtatgcaa gcattgcatct caattagtca 3300  
 gcaaccatag tcccgccctt aactccggcc atcccgcccc taactccggc cagttccggc 3360  
 cattctccgc cccatggctg actaattttt tttattttatg cagaggccga ggccgcctct 3420  
 gcctctgagc tattccagaa gtagtgagga ggcttttttg gaggcctagg cttttgcaaa 3480  
 aagctcccg gagcttgat atccattttt ggatctgac agcacgtgat gaaaaagcct 3540  
 gaactcaccg cgacgtctgt cgagaagttt ctgatcgaaa agttcgacag cgtctccgac 3600  
 ctgatgcagc tctcgagggt cgaagaatct cgtgctttca gcttcgatgt aggagggcgt 3660  
 ggatatgtcc tgcgggtaaa tagctgcgc gatggtttct acaaagatcg ttatgtttat 3720  
 cggcactttg catcgccgc gctcccgatt ccggaagtgc ttgacattgg ggaattcagc 3780  
 gagagcctga cctattgcat ctcccgccgt gcacagggtg tcacgttgca agacctgcct 3840  
 gaaaccgaac tgcccgctgt tctgcagccg gtccggagg ccattggatgc gatcgctgcg 3900  
 gccgatctta gccagacgag cgggttcggc ccattcggac cgcaaggaat cgggtcaatac 3960  
 actacatggc gtgatttcat atgcgogatt gctgatcccc atgtgtatca ctggcaaaact 4020  
 gtgatggacg acaccgtcag tgcgtccgtc ggcaggctc tcgatgagct gatgcttttg 4080  
 gccgaggact gcccgaagt ccggcacctc gtgcacggg atttcggctc caacaatgtc 4140  
 ctgacggaca atggccgcat aacagcggtc attgactgga gcgaggcgat gttcggggat 4200  
 tcccaatacg aggtcgccaa catcttcttc tggaggccgt ggttggttg tatggagcag 4260  
 cagacgcgct acttcgagcg gaggcattcc gagcttgag gatcgccgc gctccgggcg 4320  
 tataatgctc gcattggtct tgaccaactc tatcagagct tggttgacgg caatttcgat 4380  
 gatgcagctt gggcgagggt tcgatgcgac gcaatcgctc gatccggagc cgggactgtc 4440  
 gggcgtagac aaatcgcccc cagaagcgcg gccgtctgga ccgatggctg tgtagaagta 4500  
 ctgcgcgata gtggaaaccg acgccccagc actcgtccga gggcaaagga atagcacgtg 4560  
 ctacgagatt tcgattccac cgccgccttc tatgaaaggt tgggcttcgg aatcgttttc 4620

cgggacgcgc gctggatgat cctccagcgc ggggatctca tgctggagtt cttegccccac	4680
cccaacttgt ttattgcagc ttataatggg tacaaataaa gcaatagcat cacaaatttc	4740
acaaataaag catttttttc actgcattct agttgtgggt tgtccaaact catcaatgta	4800
tcttatcatg tctgtatacc gtcgacctct agctagagct tggcgtaatc atggtcatag	4860
ctgtttcctg tgtgaaattg ttatccgctc acaattccac acaacatacg agccggaagc	4920
ataaagtgtg aagcctgggg tgcctaataga gtgagctaac tcacattaat tgcgttgccg	4980
tcaactgccg ctttccagtc gggaaacctg tcgtgccagc tgcattaatg aatcggccaa	5040
cgcgcgggga gaggcgggtt gcgtattggg cgctcttcgc cttcctcgct cactgactcg	5100
ctgcgctcgg tcgttcggct gcggcgagcg gtatcagctc actcaaaggc ggtaatacgg	5160
ttatccacag aatcagggga taacgcagga aagaacatgt gagcaaaagg ccagcaaaag	5220
gccaggaacc gtaaaaaggc cgcgttgctg gcgtttttcc ataggctcgc ccccoctgac	5280
gagcatcaca aaaatcgacg ctcaagtcag aggtggcgaa acccgacagg actataaaga	5340
taccaggcgt ttccccctgg aagctccctc gtgcgctctc ctgttccgac cctgcogett	5400
accggatacc tgtccgcctt tctcccttcg ggaagcgtgg cgttttctca tagctcacgc	5460
tgtaggatc tcagttcggg taggtcgtt cgctccaagc tgggctgtgt gcacgaaacc	5520
cccgttcagc ccgaccgctg gccttatcc ggtaactatc gtcttgagtc caaccggta	5580
agacacgact tatcgccact ggcagcagcc actggtaaca ggattagcag agcgaggat	5640
gtaggcgggt ctacagagtt cttgaagtgg tggcctaact acggctacac tagaagaaca	5700
gtatttggtg tctgcgctct gctgaagcca gttaccttcg gaaaaagagt tggtagctct	5760
tgatccggca aacaaaccac cgtcggtagc ggtttttttg tttgcaagca gcagattacg	5820
cgcagaaaaa aaggatctca agaagatcct ttgatctttt ctacgggggc tgacgctcag	5880
tggaacgaaa actcacgtta agggattttg gtcattgagat tatcaaaaag gatcttcacc	5940
tagatccttt taaattaaaa atgaagtttt aaatcaatct aaagtatata tgagtaaact	6000
tggctctgaca gttaccaatg cttaatcagt gaggcacctc tctcagcgat ctgtctatct	6060
cgttcaccca tagttgcctg actccccgtc gtgtagataa ctacgatacg ggagggctta	6120
ccatctggcc ccagtctgc aatgataccg cgagaccac gctcaccggc tccagattta	6180
tcagcaataa accagccagc cggaagggcc gagcgcagaa gtggtcctgc aactttatcc	6240
gcctccatcc agtctattaa ttgttgccgg gaagctagag taagtagttc gccagttaat	6300

```

agtttgcgca acgttgttgc cattgctaca ggcacgtgg tgtcacgctc gtcgtttggt 6360
atggcttcat tcagctccgg ttcccaacga tcaaggcgag ttacatgac ccccatgttg 6420
tgcaaaaaag cggtagctc cttcggtcct ccgacgttg tcagaagtaa gttggccgca 6480
gtgttatcac tcatggttat ggcagcactg cataattctc ttactgtcat gccatccgta 6540
agatgctttt ctgtgactgg tgagtactca accaagtcac totgagaata gtgtatgcgg 6600
cgaccgagtt gctcttgccc ggcgtcaata cgggataata ccgcgccaca tagcagaact 6660
ttaaaagtgc tcatcattgg aaaacgttct tcggggcgaa aactctcaag gatcttaccg 6720
ctgttgagat ccagttcgat gtaaccact cgtgcacca actgatcttc agcatctttt 6780
actttcacca gcgtttctgg gtgagcaaaa acaggaaggc aaaatgccgc aaaaaaggga 6840
ataagggcga cacggaaatg ttgaatactc atactcttcc tttttcaata ttattgaagc 6900
atttatcagg gttattgtct catgagcgga tacatatttg aatgtattta gaaaaataaa 6960
caaatagggg ttccgcgcac atttccccga aaagtgccac ctgacgtc 7008

```

<210> 93

<211> 11693

<212> DNA

<213> Artificial

<220>

<223> Description of Artificial Sequence: Expression Vector

<400> 93

```

gttgacattg attattgact agttattaat agtaatcaat tacgggggtca ttagttcata 60
gcccatatat ggagttccgc gttacataac ttacggtaaa tggcccgccct ggctgaccgc 120
ccaacgaccc ccgccattg acgtcaataa tgacgtatgt toccatagta acgccaatag 180
ggactttcca ttgacgtcaa tgggtggagt atttacggta aactgccac ttggcagtac 240
atcaagtgtg tcatatgccg agtccgcccc ctattgacgt caatgacggg aaatggcccc 300
cctggcatta tgcccagtac atgaccttac gggactttcc tacttggcag tacatctacg 360
tattagtcac cgctattacc atgggtgatgc ggttttggca gtacaccaat gggcgtggat 420
agcggtttga ctacgggga tttccaagtc tccaccccat tgacgtcaat gggagtttgt 480
tttggcacca aaatcaacgg gactttccaa aatgtcgtaa taaccccgcc ccgttgacgc 540
aaatgggcgg taggcgtgta cgggtggagg tctatataag cagagctcgt ttagtgaacc 600
gtaagctttc ggcgcgccac ggtaccatgg gatccgaaga cgccaaaaac ataaagaaag 660
gcccggcgcc attctatcct ctagaggatg gaaccgctgg agagcaactg cataaggcta 720

```



tgaagagata cgccctggtt cctggaacaa ttgcttttac agatgcacat atcgagggtga	780
acatcacgta cgcggaatac ttcgaaatgt ccgttcggtt ggcagaagct atgaaacgat	840
atgggctgaa taaaaatcac agaatcgctg tatgcagtga aaactctctt caattcttta	900
tgccggtggt gggcgcggtt ttatcggag ttgcagttgc gcccgcgaaac gacatttata	960
atgaacgtga attgctcaac agtatgaaca ttctgcagcc taccgtagtg tttgtttcca	1020
aaaaggggtt gcaaaaaatt ttgaacgtgc aaaaaaatt accaataatc cagaaaatta	1080
ttatcatgga ttctaaaacg gattaccagg gatttcagtc gatgtacacg ttcgtcacat	1140
ctcatctacc tcccggtttt aatgaatacg attttgtacc agagtccttt gatcgtgaca	1200
aaacaattgc actgataatg aattcctctg gatctactgg gttacctaag ggtgtggccc	1260
ttccgcatag aactgcctgc gtcagattct cgcattgccag agatcctatt tttggcaatc	1320
aaatcattcc ggatactgcg attttaagtg ttgttccatt ccatcacggt tttggaatgt	1380
ttactacact cggatatttg atatgtggat ttcgagtcgt cttaatgtat agatttgaag	1440
aagagctggt tttacgatcc cttcaggatt acaaaattca aagtgcgttg ctagtaccaa	1500
ccctattttc attcttcgcc aaaagcactc tgattgacaa atacgattta tctaatttac	1560
acgaaaattgc ttctgggggc gcacctcttt cgaaagaagt cggggaagcg gttgcaaac	1620
gcttccatct tccagggata cgacaaggat atgggctcac tgagactaca tcagctattc	1680
tgattacacc cgagggggat gataaacggg gcgcggtcgg taaagttggt ccattttttg	1740
aagcgaaggt tgtggatctg gataccggga aaacgctggg cgttaatcag agaggcgaat	1800
tatgtgtcag aggacctatg attatgtccg gttatgtaaa caatccggaa gcgaccaacg	1860
ccttgattga caaggatgga tggtacatt ctggagacat agcttactgg gacgaagacg	1920
aacacttctt catagttgac cgcttgaagt ctttaattaa atacaaagga tatcagggtg	1980
cccccgctga attggaatcg atattgttac aacaccccaa catcttcgac gcgggcgtgg	2040
caggctcttc cgacgatgac gccggtgaac ttcccgccgc cgttgttggt ttggagcacg	2100
gaaagacgat gacggaaaaa gagatcgtgg attacgtcgc cagtcaagta acaaccgcga	2160
aaaagttgcg cggaggagtt gtgtttgtgg acgaagtacc gaaaggctct accggaaaaa	2220
tcgacgcaag aaaaatcaga gagatcctca taaaggccaa gaagggcgga aagtccaaat	2280
tgcgcgccg ctaactcgag aataaacaag ttaacaacaa caattgcatt cattttatgt	2340
ttcagggtca gggggagggt tgggaggttt tttaaagcaa gtaaacctc taaaaatgtg	2400

gtatggctga ttatgatccg gctgcctcgc gcgtttcggg gatgacgggtg aaaacctctg 2460  
 acacatgcag ctcccggaga cggtcacagc ttgtctgtaa gcggatgccg ggagcagaca 2520  
 agcccgtcag gcgtcagcgg gtgttgccgg gtgtcggggc gcagccatga ggtcgactct 2580  
 agaggatcga tgccccgccc cggacgaact aaacctgact acgacatctc tgccccctct 2640  
 tcgcggggca gtgcatgtaa tcccttcagt tggttggtag aacttgccaa ctgggcccctg 2700  
 ttccacatgt gacaoggggg gggaccaaac acaaaggggt tctctgactg tagttgacat 2760  
 ccttataaat ggatgtgcac atttgccaac actgagtggc ttctatctct gagcagactt 2820  
 tgcagtctgt ggactgcaac acaacattgc ctttatgtgt aactcttgcc tgaagctctt 2880  
 acaccaatgc tgggggacat gtacctccca gggggcccagg aagactacgg gaggctacac 2940  
 caacgtcaat cagagggggc tgtgtagcta ccgataagcg gacctcaag agggcattag 3000  
 caatagtgtt tataaggccc ccttggttaac cctaaacggg tagcatatgc ttcccgggta 3060  
 gtagtatata ctatccagac taacctaat tcaatagcat atgttaccca acgggaagca 3120  
 tatgctatcg aattaggggt agtaaaaggg tctaaggaa cagcgatctc tcccaccca 3180  
 tgagctgtca cggttttatt tacatggggg caggattcca cgagggtagt gaaccatttt 3240  
 agtcacaagg gcagtggctg aagatcaagg agcgggcagt gaactctcct gaatcttcgc 3300  
 ctgcttcttc attctccttc gtttagctaa tagaataact gctgagttgt gaacagtaag 3360  
 gtgtatgtga ggtgctcgaa aacaagggtt cagggtgacgc cccagaata aaatttgga 3420  
 ggggggttca gtggtggcat tgtgctatga caccaatata accctcaca accccttggg 3480  
 caataaatac tagttagga atgaaacatt ctgaatatct ttaacaatag aaatccatgg 3540  
 ggtggggaca agccgtaaag actggatgtc catctcacac gaatttatgg ctatgggcaa 3600  
 cacataatcc tagtgcaata tgatactggg gttattaaga tgtgtcccag gcagggacca 3660  
 agacagggtg accatgttgt tacactctat ttgtaacaag gggaaagaga gtggacgccg 3720  
 acagcagcgg actccactgg ttgtctctaa caccocogaa aattaaacgg ggctccacgc 3780  
 caatggggcc cataaacaaa gacaagtggc cactcttttt tttgaaattg tggagtgggg 3840  
 gcacgcgtca gccccacac gcgcacctgc ggttttgga tgtaaaataa ggggtgtaata 3900  
 acttggtgta ttgtaacccc gctaaccact gcggtcaaac cacttgccca caaaaccact 3960  
 aatggcacco cggggaatac ctgcataagt aggtggggcg gccaaagatag gggcgcgatt 4020  
 gctgcgatct ggaggacaaa ttacacacac ttgcgcctga gcgccaagca cagggttggt 4080  
 ggtcctcata ttacagaggt cgtgagagc acggtgggct aatgttgcca tgggtagcat 4140

atactaccca aatatctgga tagcatatgc tatcctaato tatatctggg tagcataggc 4200  
 taticctaato tatatctggg tagcatatgc taticctaato tatatctggg tagtatatgc 4260  
 taticctaatt tatatctggg tagcataggc taticctaato tatatctggg tagcatatgc 4320  
 taticctaato tatatctggg tagtatatgc taticctaato tgtatccggg tagcatatgc 4380  
 taticctaata gagattaggg tagtatatgc taticctaatt tatatctggg tagcatatac 4440  
 taccocaaata tctggatagc atatgctatc ctaatctata tctgggtagc atatgctatc 4500  
 ctaatctata tctgggtagc ataggctatc ctaatctata tctgggtagc atatgctatc 4560  
 ctaatctata tctgggtagt atatgctatc ctaatttata tctgggtagc ataggctatc 4620  
 ctaatctata tctgggtagc atatgctatc ctaatctata tctgggtagt atatgctatc 4680  
 ctaatctgta tccgggtagc atatgctatc ctcatgcata tacagtcagc atatgatacc 4740  
 cagtagtaga gtgggagtg c taccctttgc atatgccgc acccccaag ggggcgtgaa 4800  
 ttttcgctgc ttgtcccttt cctgctgggt gctcccatc ttaggtgaat ttaaggaggc 4860  
 caggctaaaag cgtgcgatg tctgattgct caccaggtaa atgtcgctaa tgttttccaa 4920  
 cgcgagaagg tgttgagcgc ggagctgagt gacgtgacaa catgggtatg cccaattgcc 4980  
 ccatgttggg aggacgaaaa tggtgacaag acagatggc agaaatacac caacagcacg 5040  
 catgatgtct actggggatt tattctttag tgcgggggaa tacaaggctt ttaatacgat 5100  
 tgagggcgtc tcctaacaag ttacatcaat cctgcccttc ctacccctca tctccatcac 5160  
 ctccctcatc tccgtcatct ccgtcatcac cctccgcggc agccccttcc accatagggtg 5220  
 gaaaccaggg aggcaaatct actccatcgt caaagctgca cacagtcacc ctgatattgc 5280  
 aggtaggagc gggctttgtc ataacaagg ccttaatcgc atccttcaaa acctcagcaa 5340  
 atatatgagt ttgtaaaaag accatgaaat aacagacaat ggaactccctt agcggggccag 5400  
 gttgtgggccc gggtcaggg gccattccaa aggggagacg actcaatggg gtaagacgac 5460  
 attgtggaat agcaagggca gttcctcgcc ttaggttgta aagggaggtc ttactacctc 5520  
 catatacgaa cacaccggcg acccaagttc ctctgctggg agtcccttct acgtgactcc 5580  
 tagccaggag agctcttaaa ccttctgcaa tgttctcaaa tttcgggttg gaacctcctt 5640  
 gaccaogatg cttttccaaa ccacctctct tttttgcgcc ctgcctccat cacctgacc 5700  
 ccgggggtcca gtgcttgggc cttctcctgg gtcactctgog gggccctgct ctatcgctcc 5760  
 cgggggcacg tcaggctcac catctgggccc accttcttgg tggatttcaa aataatcggc 5820

ttccccata	gggtggaaaa	atggccttct	acctggaggg	ggcctgcgcg	gtggagaccc	5880
ggatgatgat	gactgactac	tgggactcct	gggcctcttt	tctccacgtc	cacgacctct	5940
ccccctggct	ctttcacgac	ttccccccct	ggctctttca	cgctctctac	cccggggggc	6000
tccactaact	cctcgacccc	ggcctccact	acctcctoga	ccccggcctc	cactgcctcc	6060
tcgaccccg	cctccacctc	ctgctcctgc	ccctcctgct	cctgccccctc	ctcctgctcc	6120
tgccccctct	gccccctctg	ctcctgcccc	tcttgcccc	cctgctcctg	ccccctctgc	6180
ccctcctgct	cctgccccctc	ctgccccctc	tcttgctcct	gccccctcctg	ccccctctcc	6240
tgctcctgct	cctcctgccc	ctcctgctcc	tgccccctct	gccccctcctg	ctcctgcccc	6300
tcttgcccc	cctgctcctg	ccccctctgc	tcttgcccc	cctgctcctg	ccccctctgc	6360
tcttgcccc	cctgccccctc	ctgccccctc	tcttgctcct	gccccctcctg	ctcctgcccc	6420
tcttgcccc	cctgccccctc	ctgctcctgc	ccctcctcct	gctcctgccc	ctcctgcccc	6480
tcttgcccc	cctcctgctc	ctgccccctc	tgccccctct	cctgctcctg	ccccctctcc	6540
tgctcctgct	cctcctgccc	ctcctgcccc	tctcctgct	cctgccccctc	ctgccccctc	6600
tctgctcct	gccccctcctc	ctgctcctgc	ccctcctgct	cctcctgccc	ctcctcctgc	6660
tctgcccc	cctcctgctc	ctgccccctc	tgccccctct	gccccctcctg	ccccctctcc	6720
tgctcctgct	cctcctcctg	ctcctgcccc	tctgctcct	gccccctccg	ctcctgctcc	6780
tgctcctgtt	ccaccgtggg	tccctttgca	gccaatgcaa	cttggaagtt	tttgggggtct	6840
ccggacacca	tctctatgtc	ttggccctga	tctgagccg	cccgggggtc	ctgggtcttcc	6900
gctcctcgt	cctcgtcctc	ttccccgtcc	tgtccatgg	ttatcacccc	ctcttctttg	6960
agggtccactg	ccgccggagc	cttctgggtcc	agatgtgtct	cccttctctc	ctaggccatt	7020
tccaggctcct	gtacctggcc	cctcgtcaga	catgattcac	actaaaagag	atcaatagac	7080
atctttatta	gacgacgtc	agtgaataca	gggagtgcag	actcctgccc	cctccaacag	7140
ccccccacc	ctcatccct	tcatggctgc	tgctagacag	atccagggtct	gaaaattccc	7200
catcctcoga	accatcctcg	tctcctcac	caattactcg	cagcccggaa	aactcccgct	7260
gaacatcctc	aagatttgcg	tctgagcct	caagccaggc	ctcaaattcc	tgtccccct	7320
ttttgctgga	cggtagggat	ggggattctc	gggaccctc	ctcttctct	tcaaggctac	7380
cagacagaga	tgctactggg	gcaacggaag	aaaagctggg	tgcggcctgt	gaggatcagc	7440
ttatcgatga	taagctgtca	aacatgagaa	ttcttgaaga	cgaaagggcc	tcgtgatacg	7500
cctattttta	taggttaatg	tcatgataat	aatggtttct	tagacgtcag	gtggcacttt	7560

tccgggaaat	gtgcgcggaa	cccctatttg	tttatttttc	taaatacatt	caaatatgta	7620
tccgctcatg	agacaataac	cctgataaat	gcttcaataa	tattgaaaaa	ggaagagtat	7680
gagtattcaa	catttcctgt	tcgcccttat	tccctttttt	gcggcatttt	gccttcctgt	7740
ttttgctcac	ccagaaacgc	tggtgaaagt	aaaagatgct	gaagatcagt	tgggtgcacg	7800
agtgggttac	atcgaactgg	atctcaacag	cggtaagatc	cttgagagtt	ttcgccccga	7860
agaacgtttt	ccaatgatga	gcacttttaa	agttctgcta	tgtggcgcg	tattatcccg	7920
tgttgacgcc	gggcaagagc	aactcggtcg	cgcatacac	tattctcaga	atgacttggt	7980
tgagtactca	ccagtcacag	aaaagcatct	tacggatggc	atgacagtaa	gagaattatg	8040
cagtgtctgc	ataaccatga	gtgataaac	tgcggccaac	ttacttctga	caacgatcgg	8100
aggaccgaag	gagctaaccg	cttttttgca	caacatgggg	gatcatgtaa	ctcgccctga	8160
tcgttgggaa	cgggagctga	atgaagccat	accaaacgac	gagcgtgaca	ccacgatgcc	8220
tgcagcaatg	gcaacaacgt	tgcgcaaact	attaactggc	gaactactta	ctctagcttc	8280
ccggcaacaa	ttaatagact	ggatggaggc	ggataaagtt	gcaggaccac	ttctgcgctc	8340
ggcccttcgg	gctggctggt	ttattgctga	taaatctgga	gccggtgagc	gtgggtctcg	8400
cggtatcatt	gcagcactgg	ggccagatgg	taagccctcc	cgtatcgtag	ttatctacac	8460
gacggggagt	caggcaacta	tggatgaacg	aatagacag	atcgctgaga	taggtgcctc	8520
actgattaag	cattggtaac	tgtcagacca	agtttactca	tatatacttt	agattgattt	8580
aaaacttcat	ttttaattta	aaaggatcta	ggtgaagatc	ctttttgata	atctcatgac	8640
caaaatccct	taacgtgagt	tttcgttcca	ctgagcgtca	gaccccgtag	aaaagatcaa	8700
aggatcttct	tgagatcctt	ttttctctcg	cgtaatctgc	tgttgcaaa	caaaaaaac	8760
accgctacca	gcgggtggtt	gtttgcogga	tcaagagcta	ccaactcttt	ttccgaaggt	8820
aactggcttc	agcagagcgc	agataccaaa	tactgtcctt	ctagtgtagc	cgtagttagg	8880
ccaccacttc	aagaactctg	tagcaacgcc	tacatacctc	gctctgctaa	tcctgttacc	8940
agtggctgct	gccagtggcg	ataagtcgtg	tcttaccggg	ttggactcaa	gacgatagtt	9000
accggataag	gcgcagcgg	cgggctgaac	ggggggttog	tgcacacagc	ccagcttgga	9060
gcgaacgacc	tacaccgaac	tgagatacct	acagcgtgag	ctatgagaaa	gcgccacgct	9120
tcccgaagg	agaaaggcgg	acaggatatcc	ggtaagcggc	agggtcggaa	caggagagcg	9180
cacgagggg	cttcagggg	gaaacgcctg	gtatctttat	agtcctgtcg	ggtttcgcca	9240

cctctgactt gagcgctgat ttttgtgatg ctcgtcaggg gggcggagcc tatggaaaaa 9300  
 cgccagcaac ggggcctttt tacggttctt ggcccttttg tggccttgaa gctgtccctg 9360  
 atggtcgtca tctacctgcc tggacagcat ggccctgcaac gcgggcatcc cgatgccgcc 9420  
 ggaagcgaga agaatacataa tggggaaggc catccagcct cgcgtcgcga acgccagcaa 9480  
 gacgtagccc agcgcgtcgg ccccgagatg cgcgcgtgc ggctgctgga gatggcggac 9540  
 gcgatggata tgttotgcc aagggttggt tgcgcattca cagttctccg caagaattga 9600  
 ttggctccaa ttcttgaggt ggtgaatccg ttagcgaggt gccgccctgc ttcattccccg 9660  
 tggcccggtg ctgcggtttg ctggcggtgt ccccggaaga aatatatttg catgtcttta 9720  
 gttctatgat gacacaaacc ccgccagcg tottgctatt ggcgaattcg aacacgcaga 9780  
 tgcagtcggg gcggcgcggt ccgaggtcca cttcgcataa taaggtagcg cgtgtggcct 9840  
 cgaacaccga gcgacctgc agcgaccgc ttaacagcgt caacagcgtg ccgcagatcc 9900  
 cggggggcaa tgagatatga aaaagcctga actcaccgcg acgtctgtcg agaagtttct 9960  
 gatcgaaaag ttcgacagcg tctccgacct gatgcagctc tcggagggcg aagaatctcg 10020  
 tgctttcagc ttcgatgtag gagggcggtg atatgtcctg cgggtaaata gctgcgccga 10080  
 tggtttctac aaagatcgtt atgtttatcg gcactttgca tcggccgcgc tcccgatcc 10140  
 ggaagtgtt gacattgggg aattcagcga gagcctgacc tattgcatct ccgcgcgtgc 10200  
 acaggggtgc acgttgcaag acctgcctga aacogaactg ccgcgtgttc tgcagccggt 10260  
 cgcgagggcc atggatgcga tcgctgcggc cgatcttagc cagacgagcg gggttcggccc 10320  
 attcggaccg caaggaatcg gtcaatacac tacatggcgt gatttcataa gcgcgattgc 10380  
 tgatccccat gtgtatcaact ggcaaactgt gatggacgac accgtcagtg cgtccgtcgc 10440  
 gcaggctctc gatgagctga tgctttgggc cgaggactgc ccgaagtcc ggcacctcgt 10500  
 gcacgcggat ttgggtccca acaatgtcct gacggacaat ggccgcataa cagcggtcct 10560  
 tgactggagc gaggcgatgt tcggggatcc ccaatacgag gtcgccaaca ttttctctg 10620  
 gaggcggtg ttggcttgta tggagcagca gacgcgtac ttcgagcgga ggcatccgga 10680  
 gcttgacgga tcgccgggc tccgggcgta tatgtccgc attggtcttg accaactcta 10740  
 tcagagcttg gttgacggca atttcgatga tgcagcttg gcgcagggtc gatgcgacgc 10800  
 aatcgctcca tccggagccg ggactgtcgg gcgtacacaa atcgcccgca gaagcgcggc 10860  
 cgtctggacc gatggctgtg tagaagtact cgcgatagtg ggaaaccgac gcccagcac 10920  
 tcgtccggat cgggagatgg gggaggctaa ctgaaacacg gaaggagaca ataccggaag 10980

gaacccgcgc tatgacggca ataaaaagac agaataaaac gcacgggtgt tgggtcgttt 11040  
 gttcataaac ggggggttcg gtcccagggc tggcactctg tcgatacccc accgagaccc 11100  
 cattggggcc aatacgcccg cgtttcttcc ttttccccac cccaccccc aagttcgggt 11160  
 gaaggcccag ggctcgagc caacgtcggg gcggcaggcc ctgccatagc cactggcccc 11220  
 gtgggttagg gacgggggtcc cccatgggga atggtttatg gttcgtgggg gttattattt 11280  
 gggcgttgcg tggggtcagg tccacgactg gactgagcag acagacccat ggtttttgga 11340  
 tggcctgggc atggaccgca tgtactggcg cgacacgaac accggggcgc tgtggctgcc 11400  
 aaacaccccc gacccccaaa aaccaccgag cggatttctg gcgtgccaaag ctagtcgacc 11460  
 aattctcatg ttgacagct tatcatcgca gatccgggca acgttggtgc cattgctgca 11520  
 ggcgagaaac tggtaggtat ggaagatcta tacattgaat caatattggc aattagccat 11580  
 attagtcatt ggttatatag cataaatcaa tattggctat tggccattgc atacgttgta 11640  
 tctatatcat aatatgtaca tttatattgg ctcatgtcca atatgacgc cat 11693

<210> 94  
 <211> 4825  
 <212> DNA  
 <213> Artificial

<220>  
 <223> Description of Artificial Sequence: Expression vector

<400> 94  
 gacggatcgg gagatctccc gatccctat ggtgcactct cagtacaatc tgctctgatg 60  
 ccgcatagtt aagccagtat ctgctccctg cttgtgtgtt ggaggtcgct gagtagtgcg 120  
 cgagcaaaat ttaagctaca acaaggcaag gcttgaccga caattgcatg aagaatctgc 180  
 ttagggtttag gcgttttgcg ctgcttcgag atgtacgggc cagatatacg cgttgacatt 240  
 gattattgac tagttattaa tagtaatcaa ttacggggtc attagttcat agcccatata 300  
 tggagttccg cgttacataa cttacggtaa atggcccgcc tggctgaccg cccaacgacc 360  
 cccgcccatt gacgtcaata atgacgtatg ttcccatagt aacgccaata gggactttcc 420  
 attgacgtca atgggtggag tattttacgg aaactgcccc cttggcagta catcaagtgt 480  
 atcatatgcc aagtacgccc cctattgacg tcaatgacgg taaatggccc gcctggcatt 540  
 atgccagta catgacctta tgggactttc ctacttgga gtacatctac gtattagtca 600  
 tcgctattac catggtgatg cggttttggc agtacatcaa tgggcgtgga tagcggtttg 660

actcacgggg atttccaagt ctccacocca ttgacgtcaa tgggagtttg ttttggcacc 720  
 aaaatcaacg ggactttcca aaatgtcgta acaactccgc cccattgacg caaatgggog 780  
 gtaggcgtgt acggtgggag gtctatataa gcagagctct ctggctaact aagctttcgg 840  
 cgcgccgagg taccatggga tccgaagacg ccaaaaacat aaagaaaggc ccggcgccat 900  
 tctatcctct agaggatgga accgctggag agcaactgca taaggctatg aagagatacg 960  
 ccctggttcc tggaacaatt gcttttacag atgcacatat cgagggtgaac atcacgtacg 1020  
 cggaatactt cgaaatgtcc gttcggttgg cagaagctat gaaacgatat gggctgaata 1080  
 caaatcacag aatcgtcgta tgcagtgaag actctcttca attctttatg ccggtgttgg 1140  
 gcgcttatt tatcggagtt gcagttgcgc ccgcgaacga catttataat gaacgtgaat 1200  
 tgctcaacag tatgaacatt tcgcagccta ccgtagtggt tgtttccaaa aaggggttgc 1260  
 aaaaaatttt gaaogtgcaa aaaaaattac caataatcca gaaaattatt atcatggatt 1320  
 ctaaaacgga ttaccaggga tttcagtcga tgtacacgtt cgtcacatct catctacctc 1380  
 ccggttttaa tgaatacgat tttgtaccag agtcccttga tcgtgacaaa acaattgcac 1440  
 tgataatgaa ttccctctgga tctactgggt tacctaaggg tgtggccctt ccgcatagaa 1500  
 ctgcctgcgt cagattctcg catgccagag atccctatctt tggcaatcaa atcattccgg 1560  
 ataactgcgat tttaagtgtt gttccattcc atcacggttt tggaatgttt actacactcg 1620  
 gatatttgat atgtggattt cgagtcgtct taatgtatag atttgaagaa gagctgtttt 1680  
 tacgatccct tcaggattac aaaattcaaa gtgcgttgct agtaccaacc ctatcttcat 1740  
 tcttcgcaa aagcactctg attgacaaat acgatttatc taatttacac gaaattgctt 1800  
 ctggggggcg acctctttcg aaagaagtcg gggaagcggg tgcaaaacgc ttccatcttc 1860  
 cagggatacg acaaggatat gggctcactg agactacatc agctattctg attacacccg 1920  
 agggggatga taaaccgggc gcggtcggta aagttgttcc attttttgaa gcgaagggtg 1980  
 tggatctgga taccgggaaa acgctgggag ttaatcagag aggcgaatta tgtgtcagag 2040  
 gacctatgat tatgtccggt tatgtaaaca atccggaagc gaccaacgcc ttgattgaca 2100  
 aggatggatg gctacattct ggagacatag cttactggga cgaagacgaa cacttcttca 2160  
 tagttgaccg cttgaagtct ttaattaaat acaaaggata tcagggtggcc ccgctgaat 2220  
 tggaatcgat attgtttaca caccocaaca tcttcgacgc gggcgtggca ggtcttcccg 2280  
 acgatgacgc cggatgaactt cccgcgcgcg ttgttgtttt ggagcacgga aagacgatga 2340  
 cggaaaaaga gatcgtggat taogtcgcca gtcaagtaac aaccgcgaaa aagttgcgcg 2400



gaggagttgt gtttgtggac gaagtaccga aaggtcttac cggaaaactc gacgcaagaa 2460  
 aaatcagaga gatcctcata aaggccaaga agggcgga aa gtocaaattg cgcggccgct 2520  
 aactcgagaa taaaatgagg aaattgcata gcattgtctg agtaggtgtc attctattct 2580  
 ggggggtggg gtggggcagg acagcaaggg ggaggattgg gaagacaata gcaggcatgc 2640  
 tggggatgcg gtgggctcta tggcttctga ggcggaaga accagctggg gctctagggg 2700  
  
 gtatccccac gcgccctgta gcggcgcat aagcgcgcg ggtgtggtgg ttacgcgcag 2760  
 cgtgaccgct acacttgcca gcgccctagc gcccgctcct ttcgctttct tcccttcctt 2820  
 tctcgccacg ttcgccggct ttccccgtca agctctaaat cgggggtccc tttagggttc 2880  
 cgatttagtg ctttacggca cctcgacccc aaaaaacttg attaggggtga tggttcacgt 2940  
 acctagaagt tcctattcgg aagttcctat tctctagaaa gtataggaaac ttcccttgcc 3000  
 aaaaagcctg aactcaccgc gacgtctgtc gagaagtttc tgatcgaaaa gttcgacagc 3060  
 gtctccgacc tgatgcagct ctgggagggc gaagaatctc gtgctttcag cttcgatgta 3120  
 ggagggcggt gatatgtcct gcgggtaaat agctgcgccg atggtttcta caaagatcgt 3180  
 tatgtttatc ggcactttgc atcgcccgcg ctcccgatcc cggaagtgtc tgacattggg 3240  
 gaattcagcg agagcctgac ctattgcata tcccgccgtg cacagggtgt cacgttgcaa 3300  
 gacctgcctg aaaccgaact gcccgctgtt ctgcagccgg tcgcgagggc catggatgcg 3360  
 atcgctgagg ccgatcttag ccagacgagc gggttcggcc cattcggacc gcaaggaatc 3420  
 ggtcaatata ctacatggcg tgatttcata tgcgogattg ctgatcccca tgtgtatcac 3480  
 tggcaaaactg tgatggacga caccgtcagt gcgtccgtcg cgcaggctct cgatgagctg 3540  
 atgctttggg ccgaggactg ccccgaaagtc cggcacctcg tgcagcaaac aaaccaccgc 3600  
 tggtagcggg ttttttgttt gcaagcagca gattacgcgc agaaaaaaag gatctcaaga 3660  
 agatcccttg atcttttcta cggggtctga cgctcagtg aacgaaaact cacgttaagg 3720  
 gattttggtc atgagattat caaaaaggat cttcacctag atccttttaa attaaaaatg 3780  
 aagttttaaa tcaatctaaa gtatatatga gtaaacttgg tctgacagtt accaatgctt 3840  
 aatcagtgag gcacctatct cagcgatctg tctatttogg tcatccatag ttgcctgact 3900  
 ccccgctgtg tagataacta cgatacggga gggcttacca tctggcccca gtgctgcaat 3960  
 gataccgcga gaccacgct caccggctcc agatttatca gcaataaacc agccagccgg 4020  
 aagggccgag cgcagaagtg gtccctgcaac tttatccgcc tccatccagt ctattaattg 4080

```

ttgccgggaa gctagagtaa gtagttcgcc agttaatagt ttgcgcaacg ttgttgccat 4140
tgctacaggc atcgtggtgt cacgctcgtc gtttggtatg gcttcattca gctccggttc 4200
ccaacgatca aggcgagtta catgatcccc catgttggtgc aaaaaagcgg ttagctcctt 4260
cggtcctccg atcgttggtca gaagtaagtt ggccgcagtg ttatcactca tggttatggc 4320
agcaactgcat aattctctta ctgtcatgcc atccgtaaga tgcttttctg tgactggtga 4380
gtactcaacc aagtcattct gagaatagtg tatgcggcga ccgagttgct cttgcccggc 4440
gtcaatacgg gataataccg cgccacatag cagaacttta aaagtgtca tcattggaaa 4500
acgtttcttcg gggcgaaaac tctcaaggat cttaccgctg ttgagatcca gttcgatgta 4560
accactcgt gcacccaact gatcttcagc atcttttact ttcaccagcg tttctgggtg 4620
agcaaaaaaca ggaaggcaaa atgccgcaaa aaagggaata agggcgacac ggaaatgttg 4680
aatactcata ctcttctttt ttcaatatta ttgaagcatt tatcagggtt attgtctcat 4740
gagcggatac atatttgaat gtatttagaa aaataaaca ataggggttc cgcgcacatt 4800
tccccgaaaa gtgccacctg acgtc 4825
<210> 95
<211> 30
<212> DNA
<213> Artificial

<220>
<223> Synthetic Construct

<400> 95
ctgcaactcc gataaataac gcgccaaca 30

<210> 96
<211> 21
<212> DNA
<213> Artificial

<220>
<223> Synthetic Construct

<400> 96
cgggtaccga aaggtcttac c 21

<210> 97
<211> 22
<212> DNA
<213> Artificial

<220>
<223> Synthetic Construct

```

<400> 97  
 ttcttcatag ttgaccgctt ga 22

<210> 98  
 <211> 19  
 <212> DNA  
 <213> Artificial

<220>  
 <223> Synthetic Construct

<400> 98  
 gtcacgcgtcg ggaagacct 19

<210> 99  
 <211> 30  
 <212> DNA  
 <213> Artificial

<220>  
 <223> Synthetic Construct

<400> 99  
 cgatattggtt acaacaaccc aacatcttcg 30

<210> 100  
 <211> 1038  
 <212> DNA  
 <213> Artificial

<220>  
 <223> Synthetic Construct

<400> 100  
 tcgcggaggc ttggggcagc cgggtagctc ggaggtcgtg gcgctggggg ctagcaccag 60  
 cgctctgtcg ggaggcgagc cggttagggtg gaccgggtcag cggactcacc ggccagggcg 120  
 ctccggtgctg gaatttgata ttcattgata cgggttttat ccctcttctt ttttcttaaa 180  
 catttttttt taaaactgta ttgtttctcg ttttaattta tttttgcttg ccattcccca 240  
 cttgaatcgg gccgacggct tggggagatt gctctacttc cccaaatcac tgtggatttt 300  
 ggaaaccagc agaaagagga aagaggtagc aagagctcca gagagaagtc gaggaagaga 360  
 gagacggggg cagagagagc gcgcggggcg gcgagcagcg aaagcgacag gggcaaagtg 420  
 agtgacctgc ttttgggggg gaccgcccga gcgcggcggtg agccctcccc cttgggatcc 480  
 cgcagctgac cagtcgcgct gacggacaga cagacagaca ccgccccag cccagctac 540  
 cacctcctcc ccggccggcg gcggacagtg gacgcggcgg cgagccgcgg gcaggggccc 600

gagccccgcg ccggaggcgg ggtggagggg gtcggggctc gcggcgtcgc actgaaactt 660  
 ttctgtccaac ttctgggctg ttctcgcttc ggaggagccg tgggccgcgc gggggaagcc 720  
 gagccgagcg gagccgcgag aagtgctagc tcggggccggg aggagccgca gccggaggag 780  
 ggggaggagg aagaagagaa ggaagaggag agggggccgc agtggcgact cggcgctcgg 840  
 aagccgggct catggacggg tgaggcggcg gtgtgcgcag acagtgtctc agccgcgcgc 900  
 gctccccagg ccctggcccc ggcctcgggc cggggaggaa gagtagctcg ccgaggcgcc 960  
 gaggagagcg ggccgcccc cagcccgagc cggagaggga gcgcgagccg cgcgggcccc 1020  
 ggtcgggcct ccgaaacc 1038

<210> 101

<211> 1889

<212> DNA

<213> Artificial

<220>

<223> Synthetic Construct

<400> 101

gccgggcagg aggaaggagc ctccctcagg gtttcgggaa ccagatctct ctccaggaaa 60  
 gactgataca gaacgatcga tacagaaacc acgctgccgc caccacacca tcaccatcga 120  
 cagaacagtc cttaatccag aaacctgaaa tgaagggaaga ggagactctg cgcagagcac 180  
 tttgggtcog gagggcgaga ctccggcgga agcattcccc ggccgggtgac ccagcacggg 240  
 ccctcttgga attggattcg ccattttatt tttcttgctg ctaaatacc gagccgggaa 300  
 gattagagag ttttatttct gggattcctg tagacacacc caccacata catacattta 360  
 tatatatata tattatata atataaaaat aaatatctct attttatata tataaaatat 420  
 atatatctct tttttaaat aacagtgcta atgttattgg tgtcttcaact ggatgtattt 480  
 gactgctgtg gacttgagtt gggaggggaa tgttcccact cagatcctga cagggaagag 540  
 gaggagatga gagactctgg catgatcttt tttttgtccc acttgggtgg gccagggtcc 600  
 tctcccctgc ccaagaatgt gcaaggccag ggcattgggg caaatatgac ccagttttgg 660  
 gaacaccgac aaaccagcc ctggcgctga gcctctctac ccagggtcag acggacagaa 720  
 agacaaatca caggttccgg gatgaggaca ccggctctga ccaggagttt ggggagcttc 780  
 aggacattgc tgtgctttgg ggattccctc cacatgctgc acgcgcatct cgcgccaggg 840  
 ggcactgcct ggaagattca ggagcctggg cggccttcgc ttactctcac ctgcttctga 900

gttgccccagg aggccactgg cagatgtccc ggcgaagaga agagacacat tgttggaaga 960  
 agcagcccat gacagcgccc cttcctggga ctgcacctca tctcttctct gctccccctc 1020  
 ctgggggtgca gcctaaaagg acctatgtcc tcacaccatt gaaaccacta gttctgtccc 1080  
 cccaggaaac ctggttgtgt gtgtgtgagt ggttgacctt cctccatccc ctggtccttc 1140  
 ccttcccttc ccgaggcaca gagagacagg gcaggatcca cgtgcccatt gtggaggcag 1200  
 agaaaagaga aagtgtttta tatacggtag ttattttaata tcccttttta attagaaatt 1260  
 agaacagtta atttaattaa agagtagggg tttttttcag tattcttggg taatatttaa 1320  
 tttcaactat ttatgagatg tatcttttgc tctctcttgc tctcttattt gtaccggttt 1380  
 ttgtatataa aattcatgtt tccaatctct ctctccctga tcggtgacag tcactagctt 1440  
 atcttgaaca gatatttaatt tttgctaaca ctccagctctg cctccccga tccccggct 1500  
 cccagcaca cattcctttg aaagaggggt tcaatataca tctacatact atatatatat 1560  
 tgggcaactt gtatttgtgt gtatatatat atatatatgt ttatgtatat atgtgatcct 1620  
 gaaaaaataa acatcgctat totgtttttt atatgttcaa accaaacaag aaaaaataga 1680  
 gaattctaca tactaaatct ctctcctttt ttaattttta tatttggtat cattttattt 1740  
 ttggtgctac tgtttatccg taataattgt ggggaaaaga tattaacatc acgtctttgt 1800  
 ctctagtga gtttttcgag atattccgta gtacatatat atttttaaac aacgacaaag 1860  
 aaatacagat atatcttaaa aaaaaaaaaa 1889

<210> 102  
 <211> 179  
 <212> DNA  
 <213> Artificial

<220>  
 <223> Synthetic Construct

<400> 102  
 ctccctcagc aaggacagca gaggaccagc taagagggag agaagcaact acagaccccc 60  
 cctgaaaaca accctcagac gccacatccc ctgacaagct gccaggcagg ttctcttctc 120  
 ctcacatact gaccacggc tccacctct ctccctgga aaggacacca tgagcactg 179

<210> 103  
 <211> 798  
 <212> DNA  
 <213> Artificial

<220>

<223> Synthetic Construct

<400> 103

```

ggaggacgaa catccaacct tcccaaacgc ctcccctgcc ccaatccctt tattaccccc    60
tccttcagac accctcaacc tcttctggct caaaaagaga attgggggct taggggtcgga    120
acccaagctt agaactttaa gcaacaagac caccacttog aaacctggga ttcaggaatg    180
tgtggcctgc acagtgaagt gctggcaacc actaagaatt caaactgggg cctccagaac    240
tcactggggc ctacagcttt gatccctgac atctggaatc tggagaccag ggagcctttg    300
gttctggcca gaatgctgca ggacttgaga agacctcacc tagaaattga cacaagtgga    360
ccttaggcct tcctctctcc agatgtttcc agacttcctt gagacacgga gccagocct    420
ccccatggag ccagctccct ctatttatgt ttgcacttgt gattatttat tatattatta    480
ttatttatatt atttacagat gaatgtattt atttgggaga ccgggggtatc ctggggggacc    540
caatgtagga gctgccttgg ctcagacatg ttttcctgta aaacggagct gaacaatagg    600
ctgttcccat gtagccccc1ggcctctgtg ccttcttttg attatgtttt ttaaaatatt    660
tatctgatta agttgtctaa acaatgctga tttggtgacc aactgtcact cattgctgag    720
cctctgctcc ccaggggagt tgtgtctgta atcgccctac tattcagtgg cgagaaataa    780
agtttgotta gaaaagaa                                         798

```

<210> 104

<211> 7

<212> DNA

<213> Artificial

<220>

<223> Synthetic Construct

<400> 104

tatttat

7

<210> 105

<211> 33

<212> DNA

<213> Artificial

<220>

<223> Synthetic Construct

<400> 105

ttatttatta tttatttatt atttatttat tta

33

<210> 106

<211> 8  
<212> DNA  
<213> Artificial

<220>  
<223> Synthetic Construct

<400> 106  
tattttatt 8

<210> 107  
<211> 48  
<212> DNA  
<213> Artificial

<220>  
<223> Synthetic Construct

<400> 107  
taggagctgc cttggctcag acatgttttc cgtgaaaacg gagctgaa 48

<210> 108  
<211> 28  
<212> DNA  
<213> Artificial

<220>  
<223> Synthetic Construct

<400> 108  
ttttgattat gtttttttaa atatttat 28

<210> 109  
<211> 6  
<212> DNA  
<213> Artificial

<220>  
<223> Synthetic Construct

<400> 109  
aataaa 6

<210> 110  
<211> 296  
<212> DNA  
<213> Artificial

<220>  
<223> Synthetic Construct

<400> 110  
cgagcttggc tgcttctggg gcctgtgtgg ccctgtgtgt cggaaagatg gagcaagaag 60

ccgagcccca gggggcgccg cgacccctct gaccgagatc ctgctgcttt cgcagccagg 120  
 agcaccgtcc ctccccggat tagtgcgtao gagcgcccag tgccctggcc cggagagtgg 180  
 aatgatcccc gagggccagg gcgtcgtgct tccgcgcgcc ccgtgaagga aactggggag 240  
 tcttgaggga ccccgactc caagcgcgaa aaccccgat ggtgaggagc aggcaa 296

<210> 111  
 <211> 150  
 <212> DNA  
 <213> Artificial

<220>  
 <223> Synthetic Construct

<400> 111  
 aattctcgag ctogtcgacc ggtcgacgag ctcgagggtc gacgagctcg agggcgcgcg 60  
 cccggccccc acccctcgca gcaccccgcg ccccgcgccc tcccagccgg gtccagccgg 120  
 agccatgggg ccggagccgc agtgagcacc 150

<210> 112  
 <211> 21  
 <212> DNA  
 <213> Artificial

<220>  
 <223> Synthetic Construct

<400> 112  
 atggggccgg agccgcagtg a 21

<210> 113  
 <211> 612  
 <212> DNA  
 <213> Artificial

<220>  
 <223> Synthetic Construct

<400> 113  
 accagaaggc caagtccgca gaagccctga tgtgtcctca gggagcaggg aaggcctgac 60  
 ttctgtggc atcaagaggt gggagggccc tccgaccact tccaggggaa cctgccatgc 120  
 caggaacctg tctaaggaa ccttccttcc tgcttgagtt cccagatggc tggaaggggt 180  
 ccagcctcgt tggaagagga acagcactgg ggagtctttg tggattctga ggccctgccc 240  
 aatgagactc tagggccag tggatgccac agcccagctt ggccctttcc ttccagatcc 300



tgggtactga aagccttagg gaagctggcc tgagagggga agcggcccta agggagtgtc 360  
 taagaacaaa agcgacccat tcagagactg tccctgaaac ctagtactgc ccccatgag 420  
 gaaggaacag caatggtgtc agtatccagg ctttgtacag agtgcttttc tgtttagttt 480  
 ttactttttt tgttttgttt ttttaaagac gaaataaaga cccaggggag aatgggtgtt 540  
 gtatggggag gcaagtgtgg ggggtccttc tccacacca ctttgtocat ttgcaaatat 600  
 attttgaaa ac 612

<210> 114  
 <211> 336  
 <212> DNA  
 <213> Artificial

<220>  
 <223> Synthetic Construct

<400> 114  
 tcgcgaggc ttggggcagc cgggtagctc ggaggtcgtg gcgctggggg ctagcaccag 60  
 cgtctgtctg ggaggcgcag cggttaggtg gaccggtcag cggactcacc ggccagggcg 120  
 ctcggtgctg gaatttgata ttcattgatc cgggttttat ccctcttctt ttttcttaaa 180  
 cttttttttt taaaactgta ttgtttctcg ttttaattta tttttgcttg ccattcccca 240  
 cttgaatcgg gccgacggct tggggagatt gctctacttc cccaaatcac tgtggatttt 300  
 ggaaaccagc agaaagagga aagaggtagc aagagc 336

<210> 115  
 <211> 476  
 <212> DNA  
 <213> Artificial

<220>  
 <223> Synthetic Construct

<400> 115  
 tcgcgaggc ttggggcagc cgggtagctc ggaggtcgtg gcgctggggg ctagcaccag 60  
 cgtctgtctg ggaggcgcag cggttaggtg gaccggtcag cggactcacc ggccagggcg 120  
 ctcggtgctg gaatttgata ttcattgatc cgggttttat ccctcttctt ttttcttaaa 180  
 cttttttttt taaaactgta ttgtttctcg ttttaattta tttttgcttg ccattcccca 240  
 cttgaatcgg gccgacggct tggggagatt gctctacttc cccaaatcac tgtggatttt 300  
 ggaaaccagc agaaagagga aagaggtagc aagagctcca gagagaagtc gaggaagaga 360  
 gagacggggt cagagagagc gcgcgggcgt gcgagcagcg aaagcgacag gggcaaagtg 420

agtgacctgc ttttgggggt gaccgcegga ggcgcggcgtg agccctcccc cttggg 476

<210> 116  
 <211> 73  
 <212> DNA  
 <213> Artificial

<220>  
 <223> Synthetic Construct

<400> 116  
 cttttctggt tagtttttac tttttttggt ttgttttttt aaagacgaaa taaagaccca 60  
 ggggagaatg ggt 73

<210> 117  
 <211> 81  
 <212> DNA  
 <213> Artificial

<220>  
 <223> Synthetic Construct

<400> 117  
 agagaaccca ctgcttactg gcttatcgaa attaatacga ctactatag ggagacccaa 60  
 gctggctagc gtttaaactt a 81

<210> 118  
 <211> 134  
 <212> DNA  
 <213> Artificial

<220>  
 <223> Synthetic Construct

<400> 118  
 ctcgagtcta gagggcccgt ttaaaccgc tgatcagcct cgactgtggc cttctagttg 60  
 ccagccatct gttgttgtec cctccccgt cccttccttg accctggaag gtgccactcc 120  
 cactgtcctt tcct 134

# INTERNATIONAL SEARCH REPORT

International application No.

PCT/US04/26309

## A. CLASSIFICATION OF SUBJECT MATTER

IPC(7) : C12Q 1/70; C12Q 1/68; C12N 15/63

US CL : 435/6, 320.1

According to International Patent Classification (IPC) or to both national classification and IPC

## B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)

U.S. : 435/6, 320.1

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)  
Please See Continuation Sheet

## C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category *	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
Y	US 6,448,007 B1 (GIORDANO et al.) 10 September 2002 (10.09.2002), see entire document.	1-24, 41, 43-49
---		
A		31-35, 37-40, 42, 50-54
Y	US 5,859,227 A (GIORDANO et al.) 12 January 1999 (12.01.1999), see entire document.	1-24
Y	ISMAEL et al. Split-intron retroviral vectors: enhanced expression with improved safety. J Virol., March 2000, Vol. 74, No. 5, pages 2365-2371.	1-24
A	US 6,465,176 B1 (GIORDANO et al.) 15 October 2002 (15.10.2002), see entire document.	31-35, 37-54
A	US 5,928,888 A (WHITNEY) 27 July 1999 (27.07.1999), see entire document.	31-35, 37-54

☐ Further documents are listed in the continuation of Box C.

☐ See patent family annex.

\* Special categories of cited documents:

"A" document defining the general state of the art which is not considered to be of particular relevance

"E" earlier application or patent published on or after the international filing date

"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

"O" document referring to an oral disclosure, use, exhibition or other means

"P" document published prior to the international filing date but later than the priority date claimed

"T"

later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X"

document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

"Y"

document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art

"&"

document member of the same patent family

Date of the actual completion of the international search

18 June 2005 (18.06.2005)

Date of mailing of the international search report

13 JUL 2005

Name and mailing address of the ISA/US

Mail Stop PCT, Attn: ISA/US  
Commissioner for Patents  
P.O. Box 1450  
Alexandria, Virginia 22313-1450

Facsimile No. (703) 305-3230

Authorized officer

Daniel M. Sullivan

Telephone No. (571) 272-1600

*F. Roberts for*

# INTERNATIONAL SEARCH REPORT

International application No.

PCT/US04/26309

## Box No. II Observations where certain claims were found unsearchable (Continuation of item 2 of first sheet)

This international search report has not been established in respect of certain claims under Article 17(2)(a) for the following reasons:

1. ☐ Claims Nos.:  
because they relate to subject matter not required to be searched by this Authority, namely:
2. ☐ Claims Nos.:  
because they relate to parts of the international application that do not comply with the prescribed requirements to such an extent that no meaningful international search can be carried out, specifically:
3. ☐ Claims Nos.:  
because they are dependent claims and are not drafted in accordance with the second and third sentences of Rule 6.4(a).

## Box No. III Observations where unity of invention is lacking (Continuation of item 3 of first sheet)

This International Searching Authority found multiple inventions in this international application, as follows:  
Please See Continuation Sheet

1. ☐ As all required additional search fees were timely paid by the applicant, this international search report covers all searchable claims.
2. ☐ As all searchable claims could be searched without effort justifying an additional fee, this Authority did not invite payment of any additional fee.
3. ☒ As only some of the required additional search fees were timely paid by the applicant, this international search report covers only those claims for which fees were paid, specifically claims Nos.: 1-24 (in part), 31-35 and 37-54
4. ☐ No required additional search fees were timely paid by the applicant. Consequently, this international search report is restricted to the invention first mentioned in the claims; it is covered by claims Nos.:

Remark on Protest

☐  
☒

The additional search fees were accompanied by the applicant's protest.

No protest accompanied the payment of additional search fees.

# INTERNATIONAL SEARCH REPORT

International application No.  
PCT/US04/26309

## BOX III. OBSERVATIONS WHERE UNITY OF INVENTION IS LACKING

This application contains the following inventions or groups of inventions which are not so linked as to form a single general inventive concept under PCT Rule 13.1. In order for all inventions to be examined, the appropriate additional examination fees must be paid.

Group I, claim(s) 1-24, drawn to A nucleic acid construct comprising a high-level mammalian expression vector and a nucleic acid sequence encoding a reporter polypeptide wherein said nucleic acid sequence encoding a reporter polypeptide is linked to an iron response element.

Group II, claim(s) 1-24, drawn to A nucleic acid construct comprising a high-level mammalian expression vector and a nucleic acid sequence encoding a reporter polypeptide wherein said nucleic acid sequence encoding a reporter polypeptide is linked to an internal ribosomal entry site.

Group III, claim(s) 1-24, drawn to A nucleic acid construct comprising a high-level mammalian expression vector and a nucleic acid sequence encoding a reporter polypeptide wherein said nucleic acid sequence encoding a reporter polypeptide is linked to an upstream open reading frame.

Group IV, claim(s) 1-24, drawn to A nucleic acid construct comprising a high-level mammalian expression vector and a nucleic acid sequence encoding a reporter polypeptide wherein said nucleic acid sequence encoding a reporter polypeptide is linked to a male specific lethal element.

Group V, claim(s) 1-24, drawn to A nucleic acid construct comprising a high-level mammalian expression vector and a nucleic acid sequence encoding a reporter polypeptide wherein said nucleic acid sequence encoding a reporter polypeptide is linked to a G-quartet element.

Group VI, claim(s) 1-24, drawn to A nucleic acid construct comprising a high-level mammalian expression vector and a nucleic acid sequence encoding a reporter polypeptide wherein said nucleic acid sequence encoding a reporter polypeptide is linked to a 5'-terminal oligopyrimidine tract.

Group VII, claim(s) 25-30, drawn to A method of making a nucleic acid construct comprising cloning a gene and a vector in said nucleic acid construct, engineering said nucleic acid construct to prevent an expressed gene product from having a UTR not found in a target gene and linking a target UTR to said gene.

Group VIII, claim(s) 31-34, 41-54, drawn to A method of screening for a compound that modulates expression of a polypeptide comprising maintaining a cell comprising a nucleic acid molecule comprising a gene encoding a reporter polypeptide flanked by a target 5' UTR and a target 31 UTR, forming a complex with the UTR and detecting the effect of a compound on the UTR-complex.

Group IX, claim(s) 35 and 37-40, drawn to A method of screening in vivo for a compound that modulates UTR-dependent expression comprising providing a cell having a high-expression constitutive promoter upstream of a target 51 UTR, said target 5' UTR upstream from a nucleic acid encoding a reporter polypeptide, said nucleic acid encoding a reporter polypeptide upstream of a 31 UTR, contacting the cell with a compound, and detecting the reporter polypeptide.

# INTERNATIONAL SEARCH REPORT

International application No.

PCT/US04/26309

Group X, claim(s) 36, drawn to A method of screening in vitro for a compound that modulates UTR-affected expression comprising providing an in vitro translation system, contacting the in vitro translation system with a compound and a nucleic acid sequence comprising a target 5' UTR, said target 5' UTR upstream from a nucleic acid encoding a reporter polypeptide, said nucleic acid encoding a reporter polypeptide upstream of a 3' UTR, and detecting said reporter polypeptide in vitro.

The inventions listed as Groups I-X do not relate to a single general inventive concept under PCT Rule 13.1 because, under PCT Rule 13.2, they lack the same or corresponding special technical features for the following reasons:

According to PCT Rule 13.2, unity of invention exists only when the shared same or corresponding technical feature is a contribution over the prior art. The inventions listed as Groups I-X do not relate to a single general inventive concept because they lack the same or corresponding special technical feature. The Groups are united by the technical feature of a nucleic acid construct comprising a high-level mammalian expression vector and a nucleic acid sequence encoding a reporter polypeptide linked to one or more target UTRs, which target UTRs include an internal ribosomal entry site. On page 7 of the specification, reporter gene is defined as any gene whose expression can be measured. Thus, the unifying technical feature reads on any high-level mammalian expression vector comprising a nucleic acid sequence encoding a gene whose expression can be measured (essentially all genes, since the expression of any gene can be measured by northern blotting) linked to an IRES. WO 98/37189 teaches a high-level mammalian expression vector comprising a nucleic acid sequence encoding a gene whose expression can be measured operably linked to an IRES. Thus, the technical feature that unites the Groups is not a contribution over the art and the claims lack a unifying special technical feature.

The special technical feature of Group I is considered to be a reporter polypeptide linked to an iron response element, which technical feature is not shared by the nucleic acid construct of the other Groups.

The special technical feature of Group II is considered to be a reporter polypeptide linked to an internal ribosomal entry site, which technical feature is not shared by the nucleic acid construct of the other Groups.

The special technical feature of Group III is considered to be a reporter polypeptide linked to an upstream open reading frame, which technical feature is not shared by the nucleic acid construct of the other Groups.

The special technical feature of Group IV is considered to be a reporter polypeptide linked to a male specific lethal element, which technical feature is not shared by the nucleic acid construct of the other Groups.

The special technical feature of Group V is considered to be a reporter polypeptide linked to a G-quartet element, which technical feature is not shared by the nucleic acid construct of the other Groups.

The special technical feature of Group VI is considered to be a reporter polypeptide linked to a 5'-terminal oligopyrimidine tract, which technical feature is not shared by the nucleic acid construct of the other Groups.

The special technical feature of Group VII is considered to be engineering said nucleic acid construct to prevent an expressed gene product from having a UTR not found in a target gene and linking a target UTR to said gene, which process steps are not comprised by the methods of Groups VIII-X.

The special technical feature of Group VIII is considered to be forming a complex with the UTR and detecting the effect of a compound on the UTR-complex, which process steps are not comprised by the methods of Groups VII, IX and X.

The special technical feature of Group IX is considered to be providing a cell having a high-expression constitutive promoter upstream of a target 5' UTR, said target 5' UTR upstream from a nucleic acid encoding a reporter polypeptide, said nucleic acid encoding a reporter polypeptide upstream of a 3' UTR, contacting the cell with a compound, and detecting the reporter polypeptide, which process steps are not comprised by the methods of Groups VII, VIII and X.

The special technical feature of Group X is considered to be providing an in vitro translation system, contacting the in vitro translation system with a compound and a nucleic acid sequence comprising a target 5' UTR, said target 5' UTR upstream from a nucleic acid encoding a reporter polypeptide, said nucleic acid encoding a reporter polypeptide upstream of a 3' UTR, and detecting said reporter polypeptide in vitro, which process steps are not comprised by the methods of Groups VII-IX.

Accordingly, Groups I-X are not so linked by the same or corresponding special technical feature as to for a single general inventive concept.

# INTERNATIONAL SEARCH REPORT

International application No.  
PCT/US04/26309

Continuation of B. FIELDS SEARCHED Item 3:  
APS (EAST); STN (MEDLINE BIOSIS CAPLUS EMBASE CANCERLIT)  
KEYWORDS: UTR, iron response element, intron screen